

(NASA-TM-X-71645) CHAIN POOLING TO MINIMIZE
PREDICTION ERROR IN SUBSET REGRESSION (NASA)
98 p HC \$4.75 CSCL 12A

N75-18004

Unclass

G3/64 11762

NASA TECHNICAL MEMORANDUM

NASA TM X-71645

NASA TM X-71645

CHAIN POOLING TO MINIMIZE PREDICTION ERROR IN SUBSET REGRESSION

by Arthur G. Holms
Lewis Research Center
Cleveland, Ohio 44135



TECHNICAL PAPER presented at
134th Annual Meeting of the American Statistical Association
St. Louis, Missouri, August 26-29, 1974

Contents

	Page
SUMMARY	1
INTRODUCTION.	1
EXISTING THEORY OF SUBSET REGRESSION.	4
Optimality Criteria	4
Small Experiments	6
Nonlinear Models.	7
Colinearities	7
Special Techniques.	7
CHAIN POOLING COUPLES WITH PRINCIPAL COMPONENTS REGRESSION.	8
Comparison of Half-Normal Plotting with Chain Pooling	8
Procedure Based on Prior Ordering	9
Deletion Under the F-Test	9
The Largest of a Set of Chi-Square Variates	10
Chain Pooling	10
Principal Components Regression and Model Deletion.	12
Transformations Back to Original Regressors	14
CHOICE OF TRUE (POPULATION) MODEL FOR SIMULATIONS	17
Simulation Type Investigations.	17
Unfavorable Population Models	17
Functional Form of Population Model	19
Influence of Parameters on Shape of Surface	20
Choice of Population Parameter Values Unfavorable to the Decision Procedure	21
Choice of Scale Parameter Values.	23
EVALUATION OF DECISION PROCEDURE.	23
DESIGNS OF SIMULATED EXPERIMENTS.	25
Number of Treatments.	25
Plans of Experiments.	25
Aliased Parameters Initially Assumed to be Nonzero.	27
Space for which Predictions will be Evaluated	27
COMPUTER PROGRAMS	28
Flow Diagram of Program POOLES.	28
Details of Program POOLES	28
Details of Program MODEL.	33
DESCRIPTION OF RESULTS.	34
Bias Errors.. . . .	34
Strategy Options.	35
Optimal Strategies.	36

	Page
CONCLUDING REMARKS.	38
APPENDIX A - SYMBOLS.	40
APPENDIX B - PROGRAM POOLES	42
APPENDIX C - ILLUSTRATIVE OUTPUT - POOLES	52
APPENDIX D - PROGRAM MODEL.	53
REFERENCES.	59

CHAIN POOLING TO MINIMIZE PREDICTION ERRORS IN SUBSET REGRESSION

by Arthur G. Holms

Lewis Research Center

SUMMARY

A technique called chain pooling had been developed for the analysis of results of two-level fixed effects full or fractional factorial experiments not having replication. The basic strategy includes the use of one nominal level of significance for some preliminary tests and a second nominal level of significance for the final tests. Strategies were identified having approximate optimality with respect to the probabilities of Type 1 or Type 2 errors.

The subject has been reexamined from the point of view of minimizing prediction errors in the resulting subset regression equations. The investigation consisted of Monte Carlo studies using population models having geometries chosen to represent response surface applications, as motivated by some research and development in structural materials optimization. Parameter values were chosen to be unfavorable to the decision procedures. Simulated experiments were generated by adding pseudo normally distributed random errors to population values to generate "observations." Model equations were fitted to the "observations" and the decision procedures were used to delete terms. Comparison of values predicted by the reduced models with population values enabled the identification of a deletion strategy that is approximately optimal for minimizing prediction errors.

The results are proposed for any situation giving orthogonal estimators of the model coefficients. They include the cases of orthogonal designs of experiments and orthogonalizing transformations of terms of the initial regression model.

INTRODUCTION

In the development of materials for aerospace applications, the experimentation is often expensive and time consuming. In seeking optimum processing conditions or optimum compositions, the experimentation is usually done with costly raw materials and elaborate processes. The tests involved can include creep, stress rupture, fatigue, or other environmental simulations and can be time consuming. Furthermore, the complexity of the total fabrication and testing process permits the random cumulative effect of important levels of experimental error. The efficient procedure for mitigating the effects of such error, while minimizing the costs of experimentation, has been to practice mathematical modeling of the data. Modern methods for attempting to fit the most appropriate predictive equation to such data are referred to as methods of subset regression. Prac-

tical uses of such techniques in materials development have been described in Collins, Quigg, and Dreshfield (1968), Eckert and Serafini (1968), Sandrock and Holms (1969), Eiselstein (1971), and Filippi (1974). Such techniques have been made feasible by the advent of large scale computers.

Many differing statistical decision procedures have been incorporated into the computer programs that are used to pursue the objectives of subset regression. Some of them have been described by Draper and Smith (1966) and criteria for their optimality have been discussed by Hocking (1972).

Two reasons for the plethora of methods are:

1. There is no unique and widely accepted criterion for judging which of two proposed statistical procedures is the better.

2. If a unique criterion is adopted for use in measuring the goodness of a subset regression procedure, the analytical problem becomes mathematically intractable if the object is to devise a procedure of wide applicability that will satisfy the criterion. This results in many special solutions of special cases.

In the present investigation, the inhibitions of mathematical tractability were thrown off by using Monte Carlo simulations. The method of subset regression that was investigated was essentially the method of chain pooling described by Holms and Berrettoni (1969). Whereas the method had been investigated to minimize probabilities of Type 1 or Type 2 errors, the present investigation looks for strategies that minimize the maximum prediction errors. The results are offered as being generally useful because of the general applicability of the underlying assumptions. The assumptions are as follows:

1. The model to be fitted is linear in the unknown parameters.
2. The errors of the observations are independently normally distributed random variables with a zero mean and a constant variance.
3. Orthogonal estimators are available for estimating the unknown parameters of the linear model. (This orthogonality can be the result of the design of the experiment that furnished the observations, or it can be the result of an orthogonalizing transformation of the terms of the equation.)
4. The appropriate criterion of the goodness of a subset regression procedure is the smallness of the largest of the prediction error mean squares over the points of the experiment for which the experimenter desires to make predictions.
5. There is no replication available for an estimate of the "pure" error variance.

The first four of the five preceding assumptions are fundamental to the rationale of the method; however, data will be cited for believing that the method is robust against the normality part of the second assumption. The fifth assumption merely acknowledges the possibility that an altogether different method might be preferred in the presence of pure replication.

The original investigation of chain pooling had been concerned with three sizes of experiments, namely, experiments furnishing 16, 32, or 64 observations (Holms and Berrettoni (1967)). The simulations had shown drastically reduced decision error probabilities, or the equivalent, greatly improved information efficiencies, for the larger experiments. Such results suggest that the method of analysis is relatively less critical for the larger experiments, and the methods described (Holms and Berrettoni (1967)) are therefore believed to be adequate for producing small prediction errors with experiments providing 32 or more observations.

For relatively saturated experiments that are smaller than sixteen observations, the opinion is offered that such experiments are too small to provide both (1) good estimates of model coefficients and (2) a good test statistic, in cases where random errors are large enough to call for a statistical decision procedure. In other words, experiments with less than sixteen observations should be fitted with models of a fixed size with no use of conditional modeling.

Consistent with the preceding remarks, the simulations of the present investigation were all performed with experiments containing sixteen observations in the belief that such experiments are large enough to justify the use of a statistical decision procedure, but small enough so that the precise optimization of the decision procedure would be quite beneficial.

Where m_p is the number of estimable terms deleted before the decision procedure begins (or equivalently, the residual degrees of freedom), where α_p is the nominal size of the preliminary tests, and where α_f is the nominal size of the final tests, the parameters of the chain pooling strategy are the quantities m_p , α_p , and α_f . Results of the investigation are presented as recommended values of m_p , α_p , and α_f .

The results thus provide an improved method for the mathematical modeling of the small, expensive, hard to control experiments that are typical of empirical research and development in materials optimization. The improvement lies in the fact that the statistical decision procedure is optimized to minimize prediction errors, thus maximizing the accuracy with which optimum materials compositions and processes can be identified.

EXISTING THEORY OF SUBSET REGRESSION

Optimality Criteria

The underlying population model for the system being investigated is assumed to be of unknown form and to contain an unknown number of population parameters. What is known is that there is a set of variables u_1, \dots, u_g whose values can be measured or controlled within a negligible amount of error. Another variable, called the dependent variable is assumed to be observable and to contain a random component called error where the error is independently normally distributed with mean zero and constant variance. In an abstract form, the model equation for the dependent variable is written

$$Y = \mu(u_1, \dots, u_g) + e \quad (1)$$

where Y is the dependent variable, u_1, \dots, u_g are the independent variables, μ is the unknown function, and e is the error.

The next step of the procedure is to approximate the unknown function by a polynomial in the independent variables, which will be linear in a set of parameters to be estimated. The parameters will be estimated by fitting some version of the polynomial to a collection of data consisting of t matched sets of observations of y and u_1, \dots, u_g . Whether or not the data was collected from any controlled activity, the data set will be called an experiment, and the sets of values of u_1, \dots, u_g will be called treatments.

In the absence of pathological cases, the availability of the t matched sets of observations allows the fitting of a polynomial containing t unknown parameters. Thus, conceptually, an approximation to (1) is written

$$Y = \gamma_1 f_1(u_1, \dots, u_g) + \dots + \gamma_t f_t(u_1, \dots, u_g) + e \quad (2)$$

The assumption is now introduced that through prior knowledge, or through knowledge of the structure of the experiment, the functions $f_i(u_1, \dots, u_g)$ are all established. With these $f_i(u_1, \dots, u_g)$ being known functions, let

$$x_i = f_i(u_1, \dots, u_g) \quad (3)$$

and let

$$x_1 \equiv 1$$

Equation (2) is now written

$$Y = \gamma_1 + \gamma_2 x_2 + \dots + \gamma_t x_t + e \quad (4)$$

In the preceding equations, the coordinates u_1, \dots, u_g are properly called independent variables, but because the x_i are functions of the u_1, \dots, u_g , the x_i are not mathematically independent of each other. Their special role is often identified by calling them regressor variables, or simply, regressors.

Suppose equation (4) is modified by retaining only r terms beyond the constant term (not necessarily the first r terms). With $1 \leq r \leq t$ and with appropriate changes of subscripts, equation (4) is now written

$$Y = \beta_0 + \beta_1 x_1 + \dots + \beta_r x_r + e \quad (5)$$

The usual procedure is to find estimates of the β_i by the method of least squares. Assume that this has been done and that the estimates are b_i . The resulting predictive equation is

$$\hat{Y} = b_0 + b_1 x_1 + \dots + b_r x_r \quad (6)$$

For any point u_1, \dots, u_g the associated x_i can be determined and the prediction error e_p is

$$e_p = \hat{Y} - \mu(u_1, \dots, u_g) \quad (7)$$

The mean squared prediction error can be expressed as the sum of a variance and a squared bias. Where $B(\cdot)$, $E(\cdot)$, and $V(\cdot)$ are the bias, expectation, and variance of (\cdot) , respectively, the mean squared error of prediction is:

$$\begin{aligned} E(e_p^2) &= E[\hat{Y} - E(\hat{Y}) + E(\hat{Y}) - \mu]^2 \\ &= E[\hat{Y} - E(\hat{Y})]^2 + 2E\{[\hat{Y} - E(\hat{Y})][E(\hat{Y}) - \mu]\} + [E(\hat{Y}) - \mu]^2 \\ &= V(\hat{Y}) + B^2(\hat{Y}) \end{aligned} \quad (8)$$

As is well recognized, increasing r can usually be done to reduce $B^2(\hat{Y})$. As was pointed out explicitly by Walls and Weeks (1969) such increases in r are usually accompanied by increases in $V(\hat{Y})$. Selecting r and the r terms to be retained is the central problem in subset regression; however, not all authors have chosen the minimization of $E(e_p^2)$ as the objective of their procedures.

Three popular methods of subset regression are known as (a) forward selection, (b) sequential deletion, and (c) stepwise regression. Some discussion of the three methods was given by Draper and Smith (1966). Many discussions of the relative merits of differing methods of subset regression have dealt with finding the best subset of terms for a fixed number of terms [Beale (1970)]. This seems to be an incomplete formula-

tion of the problem. The question should be: "What is the best subset of terms?" with no prior specification of the number of terms in the subset.

Many of the subset regression procedures in common use do not envision the retention of a fixed number of terms, but instead allow the number retained to be determined by a statistical decision procedure. This point of view is incorporated into the procedures described by Efroymson in the publication by Ralston and Wilf (1960), by Sidik (1972a), by Holms and Berretoni (1969), and by the works of many other authors as summarized by Draper and Smith (1966). These procedures all use tests of significance, presumably in an attempt to control some combination of the probabilities of the Type 1 and Type 2 errors.

If the model selection uses tests of significance applied to the coefficients of a fitted equation, the question arises as to what form of a loss function is appropriate to the statistical decision procedure. The methods of classical statistics often give procedures whereby an arbitrary limit may be established to control the probability of Type 1 or Type 2 errors. On the other hand, real life problems seldom provide the information for a rational quantitative specification of such particular probability levels of decision errors. In many cases a reasonable objective is that of minimizing some function of the mean square error of prediction [as defined by eq. (8)]. In such cases a statistical decision procedure might operate in the guise of significance tests, but its ultimate objective would be the minimizing of prediction errors, rather than the control or probabilities of Type 1 errors.

Small Experiments

The problem under consideration is that of fitting a model equation to an experiment that resulted in a small number of observations. Consistent with the smallness of the experiment, the assumption is also made that it contains no pure replication, so that t tests or other well established procedures are not obviously valid or applicable.

Allen (1971) referred to literature that used the residual sum of squares as a criterion for choosing regressors. He pointed out that there are at least two objections to using such a criterion:

1. If the residual sum of squares were the only criterion, then all of the regressors would be used and there would be no motivation for subset regression.
2. The residual sum of squares is not directly related to the "natural" loss function which is the mean square error of prediction.

On the other hand, the methods of Allen obtain a criterion called PRESS by predicting each observation from all the other observations. The present investigation is concerned with small saturated experiments

where each observation is more or less crucial to the estimation procedure, particularly for predictions at the set of regressor values corresponding to a given observation. The PRESS criterion is therefore believed to be inappropriate to subset regression for small saturated experiments.

Nonlinear Models

In the fitting of equations to data, a large body of knowledge has been developed for the fitting of linear models. The fitting of nonlinear models is a subject of current research outside the scope of the present investigation. (The initial use of a nonlinear model is presumably dictated by prior knowledge and such prior knowledge might dictate the use of a different model selection procedure.) The restriction to linear models is no great handicap in fitting to data over sufficiently small domains because nonlinearity of the response function can be approximated by fitting a linear combination of power functions of the independent variables.

Colinearities

The coefficients of equation (4) are usually estimated by the method of least squares which gives unbiased coefficient estimates. For any of several reasons, the data analyst may wish to set some of the coefficients equal to zero, or he may wish to adjust their values as a result of certain deficiencies in the experiment.

The existence of correlations or colinearities among the regressors constitutes a potential danger in the estimation of the parameters of equation (4). Such a situation in combination with errors in the y-values can result in estimates of the y_i having excessively large absolute values. A direct treatment of this symptom is to use methods that produce biased estimates, as in the methods of Ridge Analysis discussed by Hoerl and Kennard (1970), and also by Marquardt (1970). The methods of biased estimation have the obvious disadvantage that they provide no specific procedure for the discarding or selection of terms in equation (4). They have other disadvantages as illustrated by the simulations that were performed by Newhouse and Oman (1971). Furthermore, when backward elimination is used in subset regression, the experience is usually that each successive elimination results in smaller absolute values for the coefficients that remain. This suggests that the objectives of these biased estimation procedures might also be attained by the use of a good subset regression procedure.

Special Techniques

Suppose that an experiment has been performed giving observations

of the dependent and associated independent variables, under the circumstance that the error in y is independently normally distributed with mean zero and constant variance σ^2 and that there are multicollinearities among the independent variables. One approach to such a problem is exemplified by the book by Daniel and Wood (1971) and consists of a mixture of much plotting and tentative model fitting with occasional uses of Mallows' (1973) C_p -statistic all blended with generous infusions of the experimenter's prior knowledge in an ad hoc manner for the particular set of data.

CHAIN POOLING COUPLED WITH PRINCIPAL COMPONENTS REGRESSION

The true form of the population model and the value of the error variance are assumed to be unknown. The best procedure is deemed to consist of estimating the parameters of a linear model containing as many parameters as there are observations. An appropriate procedure is to be used to set some of the parameter estimates to zero.

Two procedures available for setting parameter estimates to zero are Half-Normal Plotting (Daniel, 1959) and Chain Pooling (Holms and Berrettoni, 1969). Similarities and differences between Half-Normal Plotting and Chain Pooling will be discussed. Both of these methods require single degree of freedom orthogonal estimates of model parameters. Such estimates are ordinarily provided by the results of two-level full and fractional factorial experiments without replication. The possibility of extending such methods to cases of correlated estimates through orthogonalizing transformations ("principal components regression") will be discussed.

Comparison of Half-Normal Plotting with Chain Pooling

In Half-Normal Plotting, the absolute magnitudes of the coefficient estimates are plotted in an ordered manner on probability paper. All data interpretations are therefore dependent on the observed ordering, rather than on a prior ordering. Chain Pooling is similarly based on the observed ordering. It uses a sequence of dependent tests of hypotheses where each test is referred to the distribution function of the largest of a set of chi-square variates.

Conditional model building is just one of the objectives of Half-Normal Plotting. Other objectives include the detection of nonuniform error variance and other departures from the usual assumptions. By comparison, the objective of Chain Pooling is only that of conditional model selection, but it attempts to achieve its objective in a manner that has been established as being approximately optimal by combining a few rational considerations with a huge amount of Monte Carlo simulations.

Procedure Based on Prior Ordering

Chain pooling uses a succession of tests of significance. Another procedure using a succession of tests of significance was proposed by Kennedy and Bancroft (1971). In contrast to the methods of Half-Normal Plotting and Chain Pooling, they use a prior specified order for the tests of significance.

Deletion Under the F-Test

Subset regression procedures often proceed as follows: A multi-dimensional polynomial is fitted to the available data. If there are g independent variables and t observations, a polynomial, truncated at t terms is written having the form:

$$Y = \beta_0 + \sum_{i=1}^g \beta_i x_i + \sum_{i=1}^g \sum_{j=1}^g \beta_{ij} x_i x_j + \sum_{i=1}^g \sum_{j=1}^g \sum_{k=1}^g \beta_{ijk} x_i x_j x_k + e + \dots + \text{similar terms up to degree } d \quad (9)$$

Thus far, the assumption is that a polynomial with t terms is to be fitted to t observations. Assume instead that r terms beyond β_0 are fitted, with $r + 1 < t$ leaving

$$m_{p0} = t - (r + 1) \quad m_{p0} > 0 \quad (10)$$

degrees of freedom for the residual sum of squares. In this case m_{p0} will be called the initial degrees of freedom for error.

Let S_{mp0} be the initial residual sum of squares with m_{p0} degrees of freedom. Assuming that the residual sum of squares is due only to error (that is, assuming zero lack of fit), the quantity S_{mp0}/σ^2 is a central chi-square variable with m_{p0} degrees of freedom. Assume that among the r terms in the initial model, one or more have zero population coefficients, and assume that one of these terms, namely the i^{th} one, is selected at random, is deleted from the model, and is pooled with the residual. Then the degrees of freedom for the residual is increased by one. Let the new residual sum of squares be $S_{mp,i}$. Under the preceding assumptions the ratio

$$F = \frac{(S_{mp,i} - S_{mp,0})/1}{S_{mp,0}/m_{p0}} \quad (11)$$

has the F-distribution with 1 and m_{p0} degrees of freedom.

If only one test were to be performed, equation (11) could provide

an α -level test of the null hypothesis $\beta_i = 0$ against the alternative hypothesis $\beta_i \neq 0$. The critical region would be the region for which the test statistic exceeds the $(1 - \alpha)$ point of the cumulative distribution function of F with 1 and m_{p0} degrees of freedom. If subsequent tests are to be performed, the question arises as to whether their sensitivity can be improved by increasing the degrees freedom of the denominator through the pooling of mean squares into the denominator corresponding to coefficients that had previously tested as nonsignificant. Such a practice would invalidate the α -level of the F -test.

The Largest of a Set of Chi-Square Variates

Assume now that $j - 1$ terms have been deleted from the model that originally contained t terms, and that R_{j-1} is the residual sum of squares corresponding to the $j - 1$ terms, and that all of these terms actually represented population parameters with zero values. Let Z_{ij} be the decrease in the regression sum of squares resulting from deleting the i th term from the regression equation, and let Z_{ij} be added to R_{j-1} to form the new residual sum of squares. Assume that all of the deleted parameter estimates were independently chosen and orthogonal. If the parameter coefficient of the i th term were zero, the sum $Z_{ij} + R_{j-1}$ would then be the sum of j independent single degree of freedom central chi-square variates. Furthermore, assume that Z_{ij} is the largest of these chi-square variates. Let

$$C_j = Z_{ij} / (R_{j-1} + Z_{ij}) \quad (12)$$

Then C_j has "Cochran's" distribution as described by Cochran (1941). An upper critical point of C_j thus provides a critical region for a test of the hypothesis

$$H_0: \beta_i = 0$$

Multiplication of C_j by j gives the U_j distribution tabulated by Holms and Berrettoni (1967).

Chain Pooling

Consider a sequential deletion procedure that operates as follows: The residual sum of squares has m_p degrees of freedom. A hypothesis test is performed against the U_j distribution where $j - 1 = m_p$ and Z_j is the smallest, over all terms in the model, of the decrease, Z_{ij} , in the regression sum of squares on deletion of the i th term.

Form the statistic

$$W_j = jZ_j / (R_{j-1} + Z_j) \quad (13)$$

Because of the violations of the assumptions of Cochran's distribution that are introduced by this procedure, $Z_j + R_{j-1}$ will not in general be the sum of j random variables independently drawn from a single $\phi\chi^2(1)$ distribution with Z_j being the largest among them. As a matter of fact, in typical regression problems, an arbitrary initial residual with m_{p0} degrees of freedom might be large enough so that with $j - 1 = m_{p0}$,

$$R_{j-1}/m_{p0} > Z_j$$

On the other hand, as the sequential deletion procedure continues, by increasing j one unit at a time with each $Z_{j+1} \geq Z_j$, the statistic W_j should become an increasingly better approximation to U_j . Accordingly, the sequential deletion procedure will use the upper $1 - \alpha$ percent points of the U_j distribution as the critical values of a decision procedure based on the statistic W_j . Furthermore, the sequential deletion procedure will use the conditional pooling of the "chain pooling" strategy of Holms and Berrettoni (1969) with components $(m_{p0}, \alpha_p, \text{ and } \alpha_f)$. In this strategy, m_{p0} is the number of degrees of freedom in the initial residual. The decision procedure then consists of sequential compound hypothesis testing. The procedure begins using the W_j statistic at significance level α_p . As long as Z_j tests nonsignificant at level α_p , it is pooled into the R_{j-1} component of the denominator of W_j in preparation for the next test in the sequential deletion procedure. With the first Z_j testing significant at level α_p , testing at level α_p and the pooling of mean squares into the denominator ceases, and testing at level α_f begins. Suppose at this point that $j = j_p$. At this point, R_{j-1} is no longer augmented with values of Z_j that are not significant at level α_f . Furthermore, the critical value of U_j remains fixed at its value for $j = j_p$ and testing of the Z_j continues until some Z_j tests as significant at level α_f . Suppose this occurs at j_f . Then the associated regression coefficient is retained in the fitted equation and the decision procedure (sequential deletion) terminates. Within the strategy $(m_{p0}, \alpha_p, \alpha_f)$, the consideration of $\alpha_f > \alpha_p$ is unreasonable, and so the values of α_p and α_f , for which U_j has been tabulated and that are to be investigated, are as listed in table XI.

The use of the U_j statistic is based on the assumption that the random observation error is independently normally distributed. As was shown by Holms and Berrettoni (1967) the operating characteristics of the chain pooling procedure remain essentially unaffected for an error with the same mean and variance but with a rectangular distribution. This result is not surprising in view of the fact that the coefficient estimates are always a sum of 28^{th} observations, for which approximate normality can be expected, according to the central limit theorem. In other words the chain pooling is regarded as being generally robust with respect to the assumption of normality.

Principal Components Regression and Model Deletion

The possibility of extending a chain-type decision procedure to correlated estimates was suggested by Kennedy and Bancroft (1971) wherein they would apply orthogonalizing transformations to the original regressors. The use of orthogonalizing transformations in regression analysis had been suggested under the name of principal components regression by Kendall (1957). Further discussion of principal components and subset regression was provided by Massy (1965). The implications of combining principal components with chain pooling will be discussed in this section. Some insights into these matters are provided by numbers called the characteristic roots or the eigenvalues of the orthogonalizing transformations. The roles they play will be more fully discussed in a subsequent section. It deals with the transformations back to the original regressors that follow after a subset regression has been performed on the orthogonalized regressors.

In the method of principal components, the regressors of equation (4) are transformed to an orthogonal set and associated eigenvalues (characteristic roots) $\omega_1, \dots, \omega_t$ are also computed. Where the coordinates along the eigen vectors are w_i , the fitted equation is

$$\hat{Y} = \hat{\alpha}_0 + \hat{\alpha}_1 w_1 + \hat{\alpha}_2 w_2 + \dots + \hat{\alpha}_{t-1} w_{t-1} \quad (14)$$

The regression sum of squares is thereby partitioned into a set of mean squares (MS_i) associated with the coefficient estimates and eigenvalues as follows (p. 70, Kendall (1957)):

$$MS_i = \omega_i \hat{\alpha}_i^2 \quad (15)$$

where ω_i is the characteristic root associated with w_i .

In the procedure of Holms and Berrettoni (1969), where α_i is the population coefficient of the i th term

$$E(MS_i) = \sigma^2 + t\alpha_i^2$$

and if λ_i is the noncentrality parameter for MS_i , then

$$\lambda_i = t\alpha_i^2/\sigma^2; E(MS_i) = \sigma^2 + \lambda_i \sigma^2 = (1 + \lambda_i)\sigma^2$$

and from equation (15)

$$E(\omega_i \hat{\alpha}_i^2) = (1 + \lambda_i)\sigma^2$$

In his illustrative example, Kendall (1957) dealt with a situation

where two of the ω_i were small and where the corresponding MS_i were small. His discussion did not answer the question of what one would conclude or do if, for example, ω_4 were small and ω_5 were large, while MS_4 were large and MS_5 were small. An approach to this question was provided by Massy (1965) who differentiated between two circumstances as follows:

"a. Delete the components that are relatively unimportant as predictors of the original independent variables (X) in the problem; i.e., the components having the smallest eigen vectors should be dropped."

or

"b. Delete the components that are relatively unimportant as predictors of the dependent variable (Y) in the problem. In this case the components having the smallest values of gamma (the correlation between the components and Y) should be dropped."

The objective of the present investigation is to minimize errors of prediction, correspondingly the point of view is not necessarily either one of Massy's (1965) criteria "a" or "b."

The predictive error of the equation resulting from a subset regression procedure depends upon both bias error and variance error. The rejection of too many terms causes excessive bias error, and the retention of too many terms can cause excessive variance error. The subset regression procedure is to be optimized by minimizing the prediction error. But by the orthogonalizing process, the total variance of the observed dependent variable is partitioned into a set of independent mean squares and the smaller of these mean squares represent contributions to the variance error while their associated terms in the orthogonalized regression equation do little to reduce the bias error. Thus a procedure such as chain pooling, which rejects terms corresponding to the smaller mean squares, could be ideal for providing a subset regression procedure that would minimize the predictive error.

The basic concept of chain pooling is that beyond the $\hat{\alpha}_0$ of (14), the coefficient estimates have been ordered in the decreasing order of their associated mean squares and that η of the estimates associated with the η smallest mean squares should be set equal to zero, while ρ of them are retained. Thus η and ρ are regarded as unknown integer parameters where

$$\rho + \eta + 1 = t \quad (16)$$

The model equation is

$$Y = \alpha_0 + \alpha_1 w_1 + \dots + \alpha_\rho w_\rho + \alpha_{\rho+1} w_{\rho+1} + \dots + \alpha_{\rho+\eta} w_{\rho+\eta} \quad (17)$$

The method of chain pooling concerned the hypothesis

$$H_a: \alpha_{p+1} = \alpha_{p+2} = \dots = \alpha_{p+\eta} = 0$$

where η was to be estimated in some manner that would minimize some combination of weighted averages of the Type 1 and Type 2 error probabilities for the decisions $\alpha_i = 0$. This optimal estimate of η was to be achieved using a decision strategy $(m_p, \alpha_p, \alpha_f)$ where the parameters of the strategy are as defined by Holms and Berrettoni (1969).

The point of view of the present investigation is that the basic chain pooling approach should be retained but that the components $(m_p, \alpha_p, \alpha_f)$ of the strategy should be optimized for another criterion, namely to minimize the expected prediction error, that is, $r = \hat{\rho}$ should be chosen to give a prediction equation

$$\hat{Y} = \hat{\alpha}_0 + \hat{\alpha}_1 w_1 + \hat{\alpha}_2 w_2 + \dots + \hat{\alpha}_r w_r \quad (18)$$

such that $E(\hat{Y} - \mu(W))^2$ is minimized for some appropriately chosen value of W where W is a vector of values of the orthogonal regressors but not necessarily a vector of values for which the experiment was performed. Instead, W should be chosen according to the domain of intended application of the fitted model.

When equation (18) has been transformed back into the coordinates of equation (4), the property of minimum prediction error achieved for Y is speculated to be retained. This speculation is based on a theorem that was proven by Kennedy and Bancroft (1971) for a decision procedure that has some similarities to chain pooling. Their analysis was concerned with a prior established order of significance testing and there is no direct connection between their distributional theory and the theory for chain pooling. What they showed was that the variance error and bias, and hence the mean square prediction error of their reduced models, is invariant under an orthogonal transformation of the regressor variables.

Transformations Back to Original Regressors

The subject of principal components analysis had been developed with a view to its application to situations in multivariate analysis where all the variables are regarded as random variables. Such a development was given by Kendall and Stuart (1966), Vol. 3, pages 285-289. The present discussion of regression analysis deals with situations where the independent variables are observed or controlled with negligible error. Nevertheless the actual values given to the independent variables may be looked upon as random variates and thus for any given experiment, many aspects of the spatial distribution of the points of the regressor variables may be studied from the point of view of principal components analysis.

A valuable set of numbers characterizing the spatial distribution of the points of the experiment is the set of eigenvalues associated with the transformation of the original regressors to an orthogonal set. The relative magnitudes of these eigenvalues are measures of how widely spread the experimental points are in the directions of the eigen vectors. Thus a severe multicollinearity in the original variables is a consequence of the spatial distribution of the points being concentrated near a multi-dimensional plane. One of the new coordinates will be normal to this plane and its associated eigenvalue will be very small (Kendall and Stuart (1966)).

Some insight into the implications of the eigenvalues can be gained from figure 1, which compares two situations with differing eigenvalues. The true (but unknown) values of the dependent variable are assumed to be given (as a function of the regressor variables x_1 and x_2) by the dashed contour lines. A strong colinearity between x_1 and x_2 is illustrated by the distribution of the points of figure 1(a) which after transformation to principal components would be evidenced by a small value of w_2 . Setting $w_2 = 0$ imposes the subsidiary condition of $x_1 = x_2$ and this is a mathematical statement of the fact that the space of the x_1 was not well spanned by the experiment. The prediction equation can be either

$$Y = b_0 + b_1 x_1$$

or

$$Y = b_0 + b_1 x_2$$

but the uses of the equation must be restricted to values of x_1 and x_2 such that

$$x_1 = x_2$$

In figure 1(b), neither w_1 nor w_2 are small, however, the true population is such that MS_2 (and therefore a_2) is not significant. Because of the insignificance, the predictive equation in the principal components is

$$Y = a_0 + a_1 w_1$$

however, the predictive equation in the original coordinates contains both of them:

$$Y = b_0 + b_1 (x_1 + x_2)$$

because the values of w_1 and w_2 have now shown that the space of the x_1 is well spanned by the experiment. Also, this fact and the insignificance of MS_2 has shown that Y does not vary with w_2 , namely Y is a

constant on any line for which $x_1 + x_2$ is a constant. The preceding discussion illustrates a typical usage of the combination of principal components analysis with subset regression analysis. One situation resulted in the two equivalent equations.

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x_1$$

or

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x_2$$

whereas the other situation resulted in the model

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 (x_1 + x_2)$$

where the values of the eigenvalues were crucial to the selection and usage of the final form of the model in the original regressors.

The general situation now appears to be as follows. If the available experiment has colinearities in the original regressor variables and does not furnish orthogonal estimates of the model parameters, the methods of principal components can be used to provide orthogonal regressors that will furnish uncorrelated estimates of the coefficients of the orthogonal regressors. This process will furnish a set of eigenvalues for the matrix product of the transpose and the design matrix of the original regressors. The relative magnitudes of these eigenvalues now present information as to how badly the original regressors are correlated, or equivalently as to how badly ill-conditioned is the matrix whose inversion would ordinarily be attempted. These eigenvalues are to be judged from the standpoint of numerical analysis or computer round-off errors or similar disciplines. The eigenvalues cannot be made the subject of a statistical decision procedure.

The elements which are made the subject of the chain pooling statistical decision procedure are the MS_i that are associated with the orthogonalized regressors. The deletions are therefore in the order of the smallness of the MS_i which from equation (15) are due either to a small ω_i (as illustrated by fig. 1(a)) or to a small $\hat{\alpha}_i$ (as illustrated by fig. 1(b)) or to both. The distinctions among these cases are vital to the construction of a final model in the original coordinates. A small value of MS_i accompanied by a small value of ω_i implies that the coordinate w_i should be set equal to zero and setting that eigen vector equal to zero imposes a single degree of freedom constraint on the space of the original regressors implying that predictions are valid only for combinations of the original regressors such that $w_i = 0$. On the other hand, a small MS_i accompanied by a large ω_i would imply that α_i is small and that the dependent variable is essentially constant along lines in the space of the original regressors for which all the orthogonalized regressors except w_i are constants. In this case the term with coeffi-

cient α_i would be deleted from equation (14), and correspondingly, the original model would be limited to one less parameter, but it could contain all of the original regressors.

The preceding discussion has been given in terms of linear transformations between the x_i and the w_i . If the transformation between the u_i and the x_i (eq. (3)) are also linear, the steps in the application of principal components regression are as already described. If on the other hand, the f_i of equation (2) are nonlinear in the u_i , the consequences for the u_i of setting some of the w_i of equation (14) equal to zero would deserve special attention.

CHOICE OF TRUE (POPULATION) MODEL FOR SIMULATIONS

Simulation Type Investigations

Optimality of model selection procedures is to be investigated. An analytical investigation could provide answers in general terms. Because of the complexity of the underlying statistical procedures, an analytical investigation will not be attempted. Instead, the investigation will be conducted by the performance of Monte Carlo simulations of experiments. Such empirical investigations have the risks that the conclusion might have severely restricted ranges of applicability. In discussions to follow, the rationale will be given for the choice of the designs of the experiments and for the choice of simulated population models so that the conclusions will have wide ranges of usefulness.

The statistical decision procedure can be viewed as a game played between nature and the statistician. The statistician's strategy is his model selection procedure and he seeks to minimize the prediction error. The strategy he should use is to be found by trial and error using computer simulations. Nature's strategy must be built into the simulations. This can be done to represent nature as an indifferent opponent in which case the empirically optimized strategy will have a sort of Bayes optimality. The simulations could otherwise be done to represent nature as an aggressive opponent who would choose the worst possible population model against the statistician's best possible procedure. Under these conditions the statistician's empirically optimized procedure could be called a security strategy as defined by Luce and Raiffa (1957).

Unfavorable Population Models

In an empirical study of decision procedures for model selection in the absence of pure replication (Holms and Berrettoni (1969)) the concern was not with prediction errors resulting from the choice of model equation, but instead was confined to the count of decision errors in accepting or rejecting terms. That investigation of decision procedures was given a security quality by optimizing the recommended procedure against an "unfavorable distribution" of the parameters (regression coefficients).

On the other hand, if the procedure is to be optimized with respect to prediction errors in a manner that is also a security strategy, two "unfavorable" conditions should be introduced, namely, (1) the distribution of the values of the parameters should be unfavorable to the decision procedure, and (2) the true shape of the function to be fitted should be unfavorable with respect to what would constitute a popular or reasonable form of the initial polynomial model. Thus the criterion advanced here is not that the fitting and selection procedure should merely find good estimates of the true parameters of an unknown finite polynomial, but rather that the fitting and selection procedure should find a good predictive equation even if the unknown function is one that cannot be represented exactly by a finite polynomial.

If the true shape of the function to be fitted is particularly unfavorable to the model selection procedure, then the empirically optimized procedure will have an approximate security optimality. But another type of optimality could be of interest. The simulations could be performed with the true shape of the function to be fitted being a type that is most likely to be encountered in practice. Then if the procedure were empirically optimized to minimize mean square errors of prediction for such a function, the resulting procedure could be labeled as an approximate maximum likelihood optimal procedure. To keep the investigation to a manageable size, and still keep it highly applicable to real life situations, conditions of the investigation will be chosen that will optimize the decision procedures against a blending of the two points of view just expressed, as follows: (1) To simulate the type of function expected to be met in practice (particularly in the practice of empirical optimum seeking) it will be an "inclined bell-shaped mountain" in several variables. Differing combinations of slope and "peakedness" can then represent the differing situations that might occur, for example, in optimum seeking experiments that begin with the method of steepest ascents and then, depending on the observed topology, go to the method of local exploration. To the extent that such a "bell-shaped mountain" is typical, this aspect of the investigation will result in decision procedures that have a maximum likelihood quality. The model that will be fitted to the results of an experiment will be a polynomial in several variables with a finite number of terms. The true function is unfavorable with respect to such a model because the true model is a transcendental function, so that an exact polynomial representation would require an infinite series of terms. Of course, in any one simulation, a polynomial model containing t terms could be fitted exactly to the t observations associated with t treatments, however, the simulation will add a random error component to the observed values of the function, so that the fitted polynomial will usually fail to predict the true function values at the coordinates of the treatments. (2) In such a situation (with the model containing a number of terms equal to the number of treatments) some of the terms might be almost wholly error and thus add to the errors of prediction. Therefore a model selection procedure that rejects terms judged to be insignificant could operate to reduce the errors of prediction. The effectiveness of such a rejection procedure will be severely tested by giving the parameters values that, relative to each

other, are unfavorable to the rejection procedure. This will be done in a manner to be described under the heading of "Choice of Population Parameter Values Unfavorable to the Decision Procedure."

Functional Form of Population Model

The population model can be given the form of a bell-shaped mountain through an adaptation of the multivariate normal density function for g independent normal random variables. Such a function is given by

$$f(x_1, \dots, x_g) = \left(\frac{1}{\sqrt{2\pi}}\right)^g \left[\prod_{k=1}^g \left(\frac{1}{\sigma_k}\right) \right] e^{-\frac{1}{2} \sum_{k=1}^g \left(\frac{x_k - \mu_k}{\sigma_k}\right)^2} \quad (19)$$

where μ_k is the population mean of x_k and σ_k is the standard deviation of x_k . On any one coordinate, the function has a maximum at $x_k = \mu_k$ and it has points of inflection at $x_k = \mu_k \pm \sigma_k$. The coordinates of the treatments consist of combinations of plus and minus ones. The population model will be made asymmetrical with respect to any plan of an experiment by locating its point of symmetry at one of the possible treatment points.

The true (population) model will be a "multidimensional inclined bell-shaped mountain" because it will be given a functional form as follows:

$$\mu_{ilm} = \theta_l \left[1 + \sum_{k=1}^g \psi_k x_{ik} + \tau_m e^{-\frac{1}{2} \sum_{k=1}^g \phi_k (x_{ik} - \delta_k)^2} \right] \quad (20)$$

The parameters of equation (20) have purposes as follows:

- μ_{ilm} dependent variable, function of the x_{ik}
- θ_l scale parameter. Because $\sigma = 1$ the coefficient of variation is proportional to the reciprocal of θ_l .
- ψ_k coefficients of first degree terms
- τ_m scale factor for exponential term

ϕ_k coefficients of exponents

δ_k location parameters of mountain peak

Other symbols of equation (20) are defined in appendix A. The plan of the experiment is a $1/2^h$ replicate of a full factorial experiment with g independent variables, and the number of treatments, t , is therefore

$$t = 2^{g-h} \quad (21)$$

The subscripts have the domains

$$i = 1, \dots, t$$

$$k = 1, \dots, g$$

$$\ell = 1, \dots, \ell_\theta$$

$$m = 1, \dots, m_\tau$$

Influence of Parameters on Shape of Surface

As shown by equation (20) if $\tau_m = 0$ the only odd degree terms are $\psi_1 x_{i1}, \dots, \psi_g x_{ig}$ and therefore the surface defined by (20) is inclined with respect to the coordinates by the values given to the coefficients ψ_1, \dots, ψ_g , and they define the approximate inclination of the surface with respect to axes through the origin if τ_m is small. These inclinations of the surface are fixed by the function

$$A_i = 1 + \sum_{k=1}^g \psi_k x_{ik} \quad (22)$$

The bell shape is provided by the exponent of the exponential component of equation (20) and the exponent is

$$C_i = \sum_{k=1}^g \phi_k (x_{ik} - \delta_k)^2 \quad (23)$$

If τ_m is large in comparison with the values of ψ_k , the function given by equation (23) will provide a local maximum for (20) when (23) takes on the value of zero. The plan of the experiment will consist of treatments having coordinates such that $x_{ik} = \pm 1$. Also, the δ_k will be given values of ± 1 . Thus the local maximum will occur at a corner point of the hypercube of the experiment, namely at the corner point for which

$x_{ik} = \delta_k$ for $k = 1, \dots, g$. Furthermore, as is well known from the literature on the normal distribution density function, the function defined by equation (23) implies that the function defined by equation (20) will have two points of inflection on each coordinate axes. If these points are named x_{ak} and x_{bk} they are located by the equations:

$$\begin{aligned} x_{ak} &= \delta_k - 1/(2\phi_k)^{1/2} \\ x_{bk} &= \delta_k + 1/(2\phi_k)^{1/2} \end{aligned} \quad (24)$$

The values of x_{ak} for $\delta_k = 1$ are given by table III.

The magnitude of the function given by (20) relative to the error variance σ^2 can be adjusted by adjusting the value of the population parameter θ_ℓ . Thus the effect on optimal estimation procedures of the relation between function mean values given by (20) and the error variance σ^2 can be investigated by assigning a fixed value to σ^2 (namely $\sigma^2 = 1$) and then investigating the effects of changes in the parameter θ_ℓ . (Thus the model has been made consistent with the usual assumption that σ^2 is constant over the space of the experiment.)

Choice of Population Parameter Values Unfavorable to the Decision Procedure

The set of values $(m_p, \alpha_p, \alpha_f)$ will be called the statisticians strategy. The set of values $\psi_k, \phi_k, \theta_\ell$, and τ_m (see eq. (20)) will be called nature's strategy. The question arises as to what constitutes a least favorable, or as an approximation, what will constitute a highly unfavorable set of values of the parameters in nature's strategy?

With respect to the decision procedure, and for the two level fractional factorial experiments, an unfavorable distribution of the parameters (coefficients of the terms of the linear model) occurs if the parameters (ρ in number) have the expectations of the order statistics of a normal distribution for a sample of size ρ . (See Holms and Berrettoni (1969).) Assume that observations are drawn with a sample of size g . Let the fractiles be q_j where $j = 1, \dots, g$. The sample fractiles are

$$q_j = j/(1 + g) \quad (25)$$

These values of q_j for $g = 4, 5, 6, 7, 8$ are shown in table I. If the sample order statistics corresponding to these fractiles lie approximately on a straight line on probability paper, then the sample is concluded to have been drawn from the distribution represented by the probability paper.

Probability paper for the half normal distribution was used for figure 2 and consists of coordinates for the fractiles, q_j , and the fractile points, ψ_j . Straight lines were drawn on these graphs connecting the origin with the point defined by $q_j = q_g$ and $\psi_j = 1$. Values along these straight lines were then read to yield the values of ψ_j shown in table II. Thus if these values of ψ_j are separate population parameters for each of which single sample values are drawn, the sample values will approximate their mean values and therefore they will approximate the order statistics of samples of size g drawn from half normal distributions. Thus the ψ_k of equation (20) are likely to result in first degree regressor coefficient estimates of the fitted equation and associated mean squares that will be declared insignificant by the decision procedure. Such decision errors have two undesirable consequences, namely, first some appropriate first degree terms are immediately lost from the fitted model and second, the denominator of the test statistic is made too large so that useful higher order coefficients are also deleted. The distribution of the ψ_k thus appears to be unfavorable even if the τ_m of equation (20) is nonzero.

The ψ_j therefore constitute a distribution of parameters unfavorable to the decision procedure based on the U_j distribution and using the strategy $(m_p, \alpha_p, \alpha_f)$. (See Holms and Berrettoni (1969) for further discussion of this kind of an unfavorable distribution of parameters.)

The particular values chosen for the ψ_k of equation (20) have been discussed in connection with the description of tables I and II. For simplicity, the values of the ϕ_k of equation (20) will be set equal to the ψ_k values of table II.

The complexity of the fitted equation needed to represent a population model depends on the surface irregularities that occur within the domain of the experiment. (A basic assumption is that good prediction accuracy of the fitted equation is to be expected only within the domain of the experiment.) The domain of the experiment is defined by its corner points and they are located at $x_k = \pm 1$. Thus the domain is defined for all k by $-1 \leq x_k \leq 1$. One type of surface irregularity that can occur within the domain of the experiment is a local maximum and as may be seen from equation (20) if τ_m is sufficiently large with respect to ψ_k , a local maximum occurs at $x_k = \delta_k$ for any k . Another type of irregularity is defined by points of inflection. As developed in the discussion of equations (24) the coordinates of the points of inflection of the bell-shaped surface defined by equation (20) are given by equations (24). Consider for example, the values listed in table III for $g = 8$. Then x_{ak} of equation (24) is within the domain of the experiment for $k = 1, 2, 3, 4, 5, 6$, and 7 but is outside for $k = 8$. (All of the points of inflection defined by x_{bk} of equations (24) are outside of the domain of the experiment.)

In view of the discussion of equation (19) the final form of the population model equation is an adjusted form of equation (20) namely it is

$$\mu_{ilm} = \theta_l \left[1 + \sum_{k=1}^g \psi_k x_{ik} + \tau_m e^{-\sum_{k=1}^g \left(\frac{x_{ik} - \delta_k}{\psi_k - 1/2} \right)^2} \right] \quad (26)$$

Equation (27) results from substituting from equations (22) and (23) in equation (20). With $\phi_k = \psi_k$ the useful computing equations are (22), (23), and (27).

$$\mu_{ilm} = \theta_l [A_i + \tau_m e^{-C_i}] \quad (27)$$

Choice of Scale Parameter Values

Nine values of θ_l will be investigated, namely, $\theta_l = 1/8, 1/4, 1/2, 1, 2, 4, 8, 16$, and 32 . Three values of τ_m will be investigated, namely $\tau_m = 0.0, 1.0$, and 2.0 . The simulations will be performed for almost all combinations of these θ_l and τ_m values.

EVALUATION OF DECISION PROCEDURE

Following the selection of terms (where some of the coefficient estimates are set equal to zero), the predicted values of the dependent variable can be efficiently computed for all of the possible combinations of the independent values, by the reverse Yates method of Duckworth (1965).

Where e_{0in} are the "observation" errors, namely the pseudo normal random numbers generated in the n^{th} simulation, the "observations" are given by

$$y_{0ilm} = \mu_{ilm} + e_{0in} \quad i = 1, 2, \dots, 2^{g-h} \quad (28)$$

After the model has been fitted and insignificant terms deleted, the difference between predicted values, y_{pilmm} , of the dependent variable for the n^{th} simulation and the population mean will be called the prediction error, and thus it is

$$e_{pilmm} = y_{pilmm} - \mu_{ilm} \quad i = 1, 2, \dots, 2^g \quad (29)$$

Over the n_e simulations, the sample mean square error of prediction for a given treatment is

$$\bar{e}_{pil\bar{m}}^2 = \frac{1}{n_e} \sum_{n=1}^{n_e} e_{pil\bar{m}n}^2 \quad (30)$$

The maximum of such errors over the treatments is

$$\bar{e}_{\max}^2 = \bar{e}_{p\bar{l}\bar{m},\max}^2 = \max_{i=1,\dots,2g} \left(\bar{e}_{pil\bar{m}}^2 \right) \quad (31)$$

The mean of the squared error over the simulations and over the points of the space of the experiment is

$$\bar{e}_{p\bar{l}\bar{m}}^2 = \frac{1}{2g} \sum_{i=1}^{2g} \bar{e}_{pil\bar{m}}^2 \quad (32)$$

Equations (31) and (32) provide two criteria for measuring the effectiveness of a strategy $(m_p, \alpha_p, \alpha_f)$. The triad of particular values of m_p , α_p , and α_f that minimizes $\bar{e}_{p\bar{l}\bar{m},\max}^2$ (as given by eq. (31)) can be called a security strategy, and the triad of values that minimizes $\bar{e}_{p\bar{l}\bar{m}}^2$ can be called an approximate Bayes strategy, if the points of the space of the experiment are assumed to be equally likely of being of interest. In either case, the absolute values of squared errors would have been the prime consideration.

An example of a situation where such definitions of error would be appropriate occurs if the experimenter seeks to maximize some predicted response, such as the strength of a material as a function of its ingredients. For such an example, the region of the space of the experiment of greatest interest would be the region in the vicinity of the maximum point, where the function would most likely have its sharpest curvatures and largest errors due to lack of fit. As was discussed in connection with equation (23), this is the point where $x_{ik} = \delta_k$. For such an example, the appropriate criterion to be minimized for the choice of a strategy would seem to be the quantity $\bar{e}_{p\bar{l}\bar{m},\max}^2$ of equation (31).

The criteria of equations (31) and (32) will be evaluated using computer simulations using (in most cases) 1000 experiments. Thus the long run mean squared error of the decision procedures will be evaluated. This leaves open the question of how badly a decision procedure might perform in individual cases. One approach to this question is to evaluate the stability of the mean squared errors observed in the simulations. Thus in addition to the criteria of equations (31) and (32), two other criteria for the effectiveness of a strategy $(m_p, \alpha_p, \alpha_f)$ will be investigated. They are concerned with the stability of the quantities defined by equations (31) and (32). The instability of these criteria can be measured by the variance of the square of the prediction error. If Y

is a random variable, the unbiased estimate of the variance of Y from a sample of size n_e is given by

$$\hat{V}(Y) = \frac{1}{n_e - 1} \left[\sum_{n=1}^{n_e} Y_n^2 - \frac{1}{n_e} \left(\sum_{n=1}^{n_e} Y_n \right)^2 \right] \quad (33)$$

The random variate of interest is the squared error of prediction, namely e_{pilmn}^2 . From equation (33) the estimate of the variance of e_{pilmn}^2 is

$$\hat{V}(e_{pilmn}^2) = \frac{1}{n_e - 1} \left[\sum_{n=1}^{n_e} (e_{pilmn}^2)^2 - \frac{1}{n_e} \left(\sum_{n=1}^{n_e} e_{pilmn}^2 \right)^2 \right] \quad (34)$$

Equation (34) gives an unbiased estimate of the variance of the squared error over n_e simulations. The maximum of this quantity over the space of the simulated experiments is defined by

$$\hat{V}(\bar{e}^2)_{\max} = \hat{V}(e_{lm}^2)_{\max} = \max_{i=1, \dots, 2^g} \left[\hat{V}(e_{pilmn}^2) \right] \quad (35)$$

The arithmetic mean of the variance of the squared error over the space of the experiments is defined by

$$\hat{V}(\bar{e}_{lm}^2) = \frac{1}{2^g} \sum_{i=1}^{2^g} \hat{V}(e_{pilmn}^2) \quad (36)$$

DESIGNS OF SIMULATED EXPERIMENTS

Number of Treatments

Although statistical decision procedures can be defined operationally for any size of sample, however small, the point of view of the present investigation is that an experiment should (1) furnish all of the information to estimate the model coefficients and (2) provide adequate test statistics for the model selection procedure. A further assumption is that to provide reasonably good estimates for both the numerators and the denominators of the test statistics, the experiment must contain at least 16 observations.

Plans of Experiments

An experiment plan will be defined as a sequence of t specified treatments involving g independent variables. The experiment will be

a $1/2^h$ fraction of the full factorial experiment. The number of treatments is therefore as given by equation (21):

$$t = 2^{g-h}$$

For the present investigation with $g = 4, 5, 6, 7$, and 8 , and using 16 treatments in each experiment, the plans (using the notation of Holms (1967)) will be generated one from the other as follows:

$g = 5, h = 1$:

$$X_5 = X_2 X_3 X_4$$

$$X_0 = X_2 X_3 X_4 X_5$$

$g = 6, h = 2$:

$$X_6 = X_1 X_3 X_4$$

$$X_0 = X_2 X_3 X_4 X_5 = X_1 X_3 X_4 X_6$$

$g = 7, h = 3$:

$$X_7 = X_1 X_2 X_4$$

$$X_0 = X_2 X_3 X_4 X_5 = X_1 X_3 X_4 X_6 = X_1 X_2 X_4 X_7$$

$g = 8, h = 4$:

$$X_8 = X_1 X_2 X_3$$

$$X_0 = X_2 X_3 X_4 X_5 = X_1 X_3 X_4 X_6 = X_1 X_2 X_4 X_7 = X_1 X_2 X_3 X_8$$

These generators are also displayed by table IV. The corresponding groups of defining contrasts are displayed by table V. The plan matrix of treatments for the full factorial experiment on four independent variables may be seen by looking at the first four columns of table VI. For the fractional replicates defined in table IV, the levels of the added independent variable for $g = 5, 6, 7, 8$ are shown by the fifth, sixth, seventh, and eighth columns of table VI. Yates' notation for treatments and their order is shown in table VII. The treatments of table VI (in Yates' notation and order) and some of the lower order elements of the aliased sets of parameters (for $g = 4, 5, 6, 7$, and 8) are shown in table VIII as derived from table V.

Aliased Parameters Initially Assumed to be Nonzero

In the fitting of models to the fractional factorial experiments, one or more of three assumptions will often be made, in choosing parameters for the initial model from aliased sets of parameters.

1. A first degree term is always preferred to a two factor interaction.

2. A lower order interaction is always preferred to a relatively higher order interaction.

3. The experimenter can label the independent variables X_1, \dots, X_g in the relative order of his belief as to their relative tendency to interact. (The design of blocked fractional factorial experiments according to prior beliefs about the parameters has been discussed by Sidik and Holms (1971) and also by Sidik (1972c) and (1973).

Aliased sets of the lower order parameters corresponding to the estimates provided by the experiments are shown in table VIII. These estimates will be associated with, and given the name of, only one of the parameters in each of the aliased sets, and all other parameters in an aliased set will be set equal to zero. The parameters to be retained for the initial model will be selected in a manner consistent with the three previously stated assumptions.

In the aliased combinations, the lowest degree parameters will be assumed to be the only possible nonzero parameters in any aliased set. Where there is more than one parameter of lowest degree, the parameter with the lowest sum of its subscripts will be assumed to be the only nonzero parameter of the aliased set. If there is more than one parameter of lowest degree with lowest sum of subscripts, that parameter will be assumed nonzero which has the lowest first subscript, or lowest sum of its first two subscripts, or lowest sum of its first three subscripts, and so forth.

Based on the preceding rules (and before the deletion of terms under the statistical decision procedures) the assumed nonzero parameters of the model equations to be initially fitted to the observations are exhibited by the equations of table IX.

Space for Which Predictions Will Be Evaluated

In the case of a fractional factorial plan, the parameter estimates are based on the observations generated from population mean values according to equation (26) for the particular treatments (particular combinations of +1 and -1 for the independent variables) as specified by the fractional factorial plan. The estimated parameters then give a prediction equation which should be usable at least for predicting values of the dependent variable corresponding to the treatments actually present

in the particular fractional factorial plan. However, the experimental space consisting of all possible combinations of +1 and -1 for the independent variables can be said to have been "spanned" by the treatments actually used in the fractional factorial plan. The question arises as to whether the predicting equation should also be used for predicting values of the dependent variable for points in the space of the experiment that are not in the fractional factorial plan. The object in laying down the procedures of the present investigation is to try to simulate those assumptions that would ordinarily be made by an experimenter. In this case the question may be rephrased as: "Does the making of predictions at points which are points of the full factorial design, but not points of the fractional design actually run, constitute an act of extrapolation (not generally allowable) or an act of interpolation (generally allowable)?" The assumption now made is that such predictions will be viewed as interpolations - that the fitted model ought to be useful in making predictions at all of the points within the space "spanned" by the experiment. Thus the equations of table IX will be used for making predictions at all possible combinations of values of +1 and -1 for the X_k and not just those values listed in table VI.

Correspondingly, although the fractional factorial plans all have $2g^h = 16$, the number of points for which predicted values will be compared with population values is always 28. Therefore for $g = 4, 5, 6, 7$, and 8, the numbers of comparisons between population values and predicted values will be 16, 32, 64, 128, and 256, respectively. For these comparisons, the predicted values can be obtained using the "reversed Yates Method" as described by Duckworth (1965).

COMPUTER PROGRAMS

Flow Diagram of Program POOLES

The main logic of the program POOLES is shown by the flow diagram of figure 3. The Yates method subroutine (section 9) as called in section 2, the computation and ordering of mean squares in section 3, and the sequential deletion procedure of section 4, are essentially the same as the similar operations of "POOLMS" (as given by Amling and Holms (1973)). "POOLES" differs from "POOLMS" mainly in sections 1, 2, 5, 6, and 7; the reason for the additional sections being that "POOLES" provides a Monte Carlo evaluation of the pooling strategies based on errors of predictions.

Details of Program POOLES

Section 1: Declarations, constants, population means, and strategy. - The constants defining the populations, the experiments, and the sequential deletion strategy, are read from data cards in the following order, with the order of the fields being the same as the order of the symbols in the following description:

Format	Description
(13A6, A2)	Arbitrary literal information such as particular use of program, date of last change, and so forth
(3I5)	LGMH, NE
(I5, 8F5.0)	KG, (PSI(K), K=1, KG)
(I5, 8F5.0)	KG, (PHI(K), K=1, KG)
(I5, 8F5.0)	KG, (DEL(K), K=1, KG)
(I5, 8F5.0)	LTH, (THETA(L), L=1, LTH)
(I5, 8F5.0)	MTAU, (TAU(M), M=1, MTAU)
(3I5)	MP, KP, KF

After the declarations and initial constants have been read, the major operation is the synthesis of the population mean values in accordance with equations (22), (23), and (27). The solutions of these equations (the synthesized values of $\mu_{i\ell m}$) are to be stored in an array coded as YMU (I,L,M). The points in the space of the experiment are accounted for by the loop "DO 20 I = 1, IP." Within that loop, construction of the A_i and C_i of equations (22) and (23) is done within the loop "DO 15 K=1, KG." The loop contains some special NASA-Lewis functions available within the IBM 7044-7094 direct couple system. As used in POOLES the Boolean function AND ((I-1), J) gives the logical intersection of two 36 bit integer arguments as a real function. As used in POOLES the Lewis shift function IARS (KML, AX) gives the integer function resulting from the accumulator right shift of the real variable AX by the integer number KML of binary places.

The constructions of $\mu_{i\ell m}$ over the number of predicted values IP are then repeated for all values $\ell = 1, \dots, \ell_g$ and $m = 1, \dots, m_t$ by the loops "DO 49 M=1, MTAU" and "DO 48 L=1, LTH." The array YMU (I,L,M) then contains all the population means as determined by equation (27).

For the given experiment plan and for the established sets of population means, more than one model deletion strategy can be evaluated. The strategy to be evaluated, and its subsequent evaluation, begins with the statement "50 READ (5,824) MP, KP, KF." On completion of the evaluation of a particular strategy, control is transferred back to statement 50. Subsequent strategy investigations are initiated by reading additional data cards according to statement 50. The operation of the program ends when such cards are exhausted.

The error simulations are generated so that each strategy is com-

pared for the same set of n_e time 28^h random numbers. This is achieved by reinitializing the random number generator for each new strategy with the statement "CALL SAND (XS)." The procedure generates a sequence of pseudo random numbers with a rectangular distribution by taking the low order 36 bits of the product $r_{r-1} * K$ where r_{r-1} = previous random number and $r_0 = 1$ and $K = 515$. This fixed point number is then floated and returned to the calling program as a floating point number between 0 and 1 (Tausky and Todd (1956)).

The prediction errors and their squares are stored in the arrays ERSQ (I,L,M) AND ERSQSQ (I,L,M). These arrays are initially cleared by the loops terminating at statements 97, 98, and 99. The arrays ERSQ and ERSQSQ used a large amount of storage. Many simulations used $L = 6$ and $M = 6$ but storage was exceeded with $I = 256$. To accommodate $I = 256$, the necessarily smaller values of L and M were used by making the appropriate changes in the DIMENSION and FORMAT statements.

Section 2: Simulations and model fitting. - The number of experiments simulated is NE. The performance of these experiments and their analysis is controlled by the loop: "DO 699 N=1, NE." Within each experiment, the pseudo-normal random numbers for the IT treatments are generated within the loop "DO 215 I=1, IT, 2," the transformation to approximate normality being that described by Box and Muller (1958). Each set of IT random numbers for an experiment is used with all possible combination of the population parameters θ_λ and τ_m through the statements "DO 690 M=1, MTAU" and "DO 680 L=1, LTH." For all of these cases, the simulated observation errors, as stored in RN(I), are to be added to the population mean values (stored in YMU (I,L,M)) for the particular treatments ($I = 1, \dots, IT$) that were specified for particular values of g by table VI. Accordingly as $g = 4, 5, 6, 7$, or 8 (that is, according to the value of KG) control is transferred to statements 204, 205, 206, 207, or 208.

After synthesizing the "observed" values of YOBS(I) the "SUBROUTINE YATES" (section 9) ending with statement 909 is used to compute the array B(I) which contains (except for division by the number of treatments) the Yates estimates of the parameters of table VIII. (The operation of the statements within SUBROUTINE YATES was described by Amling and Holms (1973).)

Section 3: Construction and ordering of mean squares. - The mean squares are formed from the parameter estimates (for those terms beyond β_0) and a pointer function is created within the loop "DO 309 I=1, IT." The strategy of doing no sequential deletion is represented by the code $m_p = 0$, and if $m_p = 0$, control is transferred to section 5. Otherwise, ordering of the mean squares is done within the loop "DO 313 J=1, ITM2."

Section 4: Deletion of terms. - The statistical decision procedure using the strategy (m_p, α_p, α_f) is applied in this section as was described for POOLMS by Amling and Holms (1973). The sequential deletion begins with the pooling of the initial residual within the loop

"DO 415 J=1, MP." The sequential deletion based on conditional pooling ("chain pooling") then proceeds within the loop "DO 419 J=MPPI, ITMI." This results in the retention in the model of β_0 and rejection from the model of all terms with mean squares as small or smaller than the ordered mean square with subscript JETA.

Section 5: Predictions. - Predicted values of the dependent variable for all the 2^g points of the experiment space spanned by the fractional factorial experiment are computed in this section using SUBROUTINE YATES and the reversed Yates method as proposed by Duckworth (1965).

The operation of Yates' method followed by the "reversed Yates method" is illustrated by the following table for a 2^2 experiment:

YATES METHOD

YOBS		B	B/FIP
Y_1	$Y_1 + Y_2$	$Y_1 + Y_2 + Y_3 + Y_4$	$(Y_1 + Y_2 + Y_3 + Y_4)/4$
Y_2	$Y_3 + Y_4$	$Y_2 - Y_1 + Y_4 - Y_3$	$(Y_2 - Y_1 + Y_4 - Y_3)/4$
Y_3	$Y_2 - Y_1$	$Y_3 + Y_4 - Y_1 - Y_2$	$(Y_3 + Y_4 - Y_1 - Y_2)/4$
Y_4	$Y_4 - Y_3$	$Y_4 - Y_3 - Y_2 + Y_1$	$(Y_4 - Y_3 - Y_2 + Y_1)/4$

REVERSED YATES METHOD

YOBS		B	YPRED
$(Y_4 - Y_3 - Y_2 + Y_1)/4$	$(Y_4 - Y_2)/2$	Y_4	Y_1
$(Y_3 + Y_4 - Y_1 - Y_2)/4$	$(Y_2 + Y_4)/2$	Y_3	Y_2
$(Y_2 - Y_1 + Y_4 - Y_3)/4$	$(Y_3 - Y_1)/2$	Y_2	Y_3
$(Y_1 + Y_2 + Y_3 + Y_4)/4$	$(Y_1 + Y_3)/2$	Y_1	Y_4

In the case of the computer program, there are 2^{g-h} parameters estimated from a fractional factorial experiment. They are exhibited by the prediction equations of table IX. The reversed Yates method requires the insertion of zeros for all parameters of a 2^g model that are not estimated in a 2^{g-h} experiment. To make the 2^g predictions corresponding to experimental points of the full factorial experiment, the 2^{g-h} coefficients together with $2^g - 2^{g-h}$ zeros are listed in the correct Yates order for the 2^g coefficients of a full factorial experiment. The statements of section 5 then combine estimates for such estimable parameters with zeros for the nonestimable parameters in the reverse Yates method. This is done according as $g = 4, 5, 6, 7$, or 8 by the transfer of control to statements 540, 550, 560, 570, or 580, respectively. This list is then reversed and the reversed Yates method is used as in section 5.

Section 6: Accumulation of errors. - The squared error for each prediction is accumulated (as required by equation (30)) in the array ERSQ(I,L,M) as computed with the loop "DO 609 I=1, IP." These accumulations are stored for each combination of L and M as indicated by the loops terminating at statements 680 and 690, and this process is repeated for each of the n_e sets of IT random numbers as indicated by the loop terminating at statement 699. For the purpose of computing the variance of the squared error of prediction, the quantity

$$\sum_{n=1}^{n_e} (e_{pilmm}^2)^2$$

of equation (34) is computed within the loop ending at statement 609 and stored as ERSQSQ(I,L,M).

Section 7: Determination of maximum and mean squared errors and their variances. - The purpose of this section is to determine maximums and means of the prediction errors over the space of the experiment after the errors have been evaluated over that space by accumulating over the simulations. The accumulation over the number, n_e , of simulations had been stored in the array ERSQ(I,L,M). For particular L, and M, the determination of the largest prediction error over the space of the experiment as defined by equation (31) is done through repeated use of the library subroutine AMAX1, which determines a real number as a function of two real arguments. This is done within the loop "DO 750 I=1, IP." The summation for the mean squared prediction error over the space of the experiment as required by equation (32) is also done within the same loop terminating at statement 750. After division by the appropriate divisors, these two evaluations of error are stored in the arrays (ERSQMX(L,M) and AVERSQ(L,M). The quantity

$$\sum_{n=1}^{n_e} (e_{pilmm}^2)^2 - \frac{1}{n_e} \left(\sum_{n=1}^{n_e} e_{pilmm}^2 \right)^2$$

is computed and stored as TEM within the loop ending at statement number 750. The quantity $\hat{V}(e_{lm}^2)_{\max}$ defined by equation (35) is determined to be the maximum of the values of TEM as determined by

$$E = \text{AMAX1}(E, \text{TEM})$$

and from this maximum, $\hat{V}(e_{lm}^2)_{\max}$ is computed and stored with the statement

$$\text{VESQMX}(L,M) = E/\text{FNEM1}$$

The sum of the values of TEM as given by

$$F = F + TEM$$

is then used to compute $\hat{V}(e_{lm}^2)$ according to equation (36) using the statement

$$AVVESQ(L,M) = F/FEMLIP$$

The computation ends if the data for MP, KP, and KF is exhausted; otherwise a new strategy is investigated by returning control to statement 50.

Details of Program MODEL

The main work of the investigation was done with program POOLES. Some properties of the population model were explored using a second program called MODEL. A summary of the two programs showing their names, principal options and outputs is contained in table X.

The program MODEL was used to exhibit the mean values of the population without any variability introduced by randomness. Thus population mean values were computed according to equation (26), but there was no random number generator to generate errors with specified properties as suggested by equation (28). The only errors are the round-off errors of the computing system, and the system consisted of an IBM 7044 monitor direct coupled to an IBM 7094 central computer. For that system, the round-off error typically occurs in the eighth significant figure of decimal output. The lack of intentional random error in the program MODEL was symbolized by showing $\sigma^2 = 0$ in table X.

The manner in which the components (m_p , α_p , α_f) influence the decision procedure was described by Holms and Berréttoni (1969). The manner in which these parameters are read into the computer program was described for a different program called POOLMS by Amling and Holms (1973). A major distinction between POOLMS and the programs of the current investigation is that the value $m_p = 0$ is now possible. Its use merely implies that no deletion of any kind will be applied to the initial model. Therefore, as suggested by the parameters of table X, the program MODEL generates no random error, and uses no statistical decision procedures on the initial fitted models (as they were given in table IX).

The output of MODEL results from the following:

1. Population mean values are generated according to equation (26) and listed as YMU(I,L,M).

2. The population mean values are used directly in the Yates procedure to estimate the parameters of the initial models of table IX. These estimates are listed as BILM(I,L,M).

3. The reversed Yates method (Duckworth (1965)) is used to compute predicted values for all of the 2^g treatments of the full factorial experiment. Differences between these predicted values and the population values from equation (26) are defined as bias and are listed as the array named $BIAS(I,L,M)$.

Because 16 parameters are estimated from orthogonal contrasts of the 16 population values for the 16 treatments of the 2^4 experiment, the bias for these treatments consists merely of computer round-off error. For the cases of $g > 4$ (namely for the fractional replicate experiments) the bias values for the points of the full 2^g experiment not included in the 16 points used for the estimating should be much larger than just round-off error.

DESCRIPTION OF RESULTS

Bias Errors

The statistical decision procedure was investigated by fitting polynomial equations to simulations of fractional factorial experiments of type 2^{g-h} where $g = 4, 5, 6, 7$, and 8 , and where $h = g - 4$. The polynomials that might reasonably be fitted in such circumstances were listed in table IX. If such fitted models are to be used for making predictions at all of the hypercube points of the corresponding 2^g , (full factorial) experiment, then bias errors can be expected for those predictions that are made at points of the space of the experiment that were not points for which observed values were recorded. In order to best understand the results of the Monte-Carlo simulations that will include both bias and variance errors, the errors due to bias alone should be investigated.

As previously described, the program MODEL was written to exhibit the bias errors that exist when there are no random errors and when there is no model deletion. Some results are shown by table XII. As is shown by table IX, the case of $g = 4$ has a starting model with 16 coefficients fitted to 16 "observations" and with $h = 0$, predictions are made only for the points of the 16 observations. The result consists of 16 predictions containing only round-off error and no bias error. Thus for $g = 4$ and $h = 0$ table XII shows the consequent zero bias for all combinations of values of θ and τ as indicated by the fact that the maximum absolute value of the bias is so listed for all combinations such that $\theta = 1, 32$, and $\tau = 0, 1, 2, 4, 8, 16$.

In the cases of the fractional factorial experiments, the population model (as shown by eq. (26)) need contain no more than nine first degree terms for $g = 5, 6, 7$, and 8 , for the case of $\tau = 0$. For this case, the fitting of the equations of table IX to the 16 "observations" from the fractional factorial experiments always results in zero bias as confirmed by the results from program MODEL in table XII.

For the cases of $h > 0$ (the fractional factorial experiments) and

for $\tau \geq 1$ (the presence of higher than first degree terms in the population model of equation (26)), the maximum absolute values of bias in table XII were always greater than zero. Furthermore the bias values were always quite large in comparison with the largest population values for all values of $\tau > 1$. (The largest bias values were always less than 20 percent of the largest population values for $\tau = 1$ but were always larger than 20 percent of the largest population values for $\tau > 1$ for the values of $\theta = 1$ and $\theta = 32$ as listed in table XII.) Based on these observations of very large bias with $\tau > 1$, the consideration of optimal decision procedure strategies (m_p , α_p , α_f) will be mainly limited to the conditions of $\tau = 0$ and $\tau = 1$.

Strategy Options

The conceptually simplest strategy is to do no deletion. Then no mean squares are initially available for a test statistic and this fact and the associated strategy will be symbolized by writing $m_p = 0$. Consistent with such a strategy, there will be no mean squares tested for pooling which is represented by $\alpha_p = 1.0$ and none can be judged insignificant which is represented by $\alpha_f = 1.0$.

The choice of just the smallest mean square as the denominator of the test statistic is represented by $m_p = 1$, and subsequent pooling and significance tests are represented by $\alpha_p \leq 1$ and $\alpha_f \leq 1$ as discussed by Holms and Berrettoni (1969).

Consider the situation where t treatments have been used to produce t observations and there are therefore t degrees of freedom to be partitioned between single degree of freedom estimates of model parameters and the degrees of freedom for the denominator of the test statistic at the start of the decision procedure. If the data analyst does not have prior information on the complexity required of the model, the wise strategy would seem to consist of starting with the maximum possible number of degrees of freedom for the model and therefore with the minimum possible number of degrees of freedom for the test statistic, namely, use $m_p = 1$. The use of $\alpha_p < 1$ will then permit the denominator of the test statistic to be augmented with additional degrees of freedom thus increasing its sensitivity. Based on this line of reasoning, no investigations with POOLLES were run with $m_p > 1$.

The combinations of α_p and α_f available for investigation are listed in table XI. Not all of these combinations were investigated with $n_e = 1000$ because some preliminary work with $n_e = 10$ had suggested that for values of $\theta > 1$ a strategy becomes ineffective if both (1) a very large value of α_p is used to severely limit the number of degrees of freedom for the test statistic and if (2) this is followed with a very stringent final test of significance as called for by a very small value of α_f . Consistent with this concept, the values of α_f were mainly limited to $\alpha_f \geq \alpha_p/10$ with the exceptions occurring for $\theta \leq 1$.

Strategies for which α_p and α_f are both large will be called aggressive strategies. They tend to retain most of the original terms in the model. Strategies for which α_p and α_f are both small will be called defensive strategies. They have strong tendencies to discard terms. The intermediate strategies where α_p is large and α_f is small will be called moderate strategies. They tend to pool a relatively small number of terms into the denominator of the test statistic and because of the ordering of the terms by absolute magnitudes, the denominator remains fairly small, and so the final test at level α_f is a relatively sensitive test, and not as defensive as it would be if α_p had been small for the same α_f .

Optimal Strategies

Illustrative output from POOLLES is shown by appendix C. Similar outputs for many strategies (m_p , α_p , α_f) were scanned to find the strategy giving the smallest value of \bar{e}_{\max}^2 (eq. (31)) for given values of g , θ , and τ . Results are shown by table XIII together with associated values of $V(\bar{e}_{\max}^2)$ (eq. (35)). The resulting values of \bar{e}_{\max}^2 for large θ and $\tau = 0$ can be compared with expected values. If the terms in the population model are much larger than the variance error, and if in the population model (eq. (26)), $\tau = 0$, then the number of terms that should be retained in the population model is $g + 1$. With $\sigma^2 = 1$ and with $g + 1$ terms estimated from 16 observations the total expectation of error variance for the prediction based on $g + 1$ independently estimated terms is

$$E(\bar{e}^2) = (g + 1)/16$$

From the preceding equation and from table XIII, the values of $E(\bar{e}^2)$ and \bar{e}_{\max}^2 for $\theta = 32$, $\tau = 0$, for the several values of g were as follows:

g	$(g + 1)/16$	\bar{e}_{\max}^2 ($\theta = 32$, $\tau = 0$)
4	0.3125	0.3447
5	.375	.4177
6	.4375	.5038
7	.500	.5745
8	.5625	.6793

Some of the optimal strategies of table XIII for $g > 4$ are seen to be quite defensive for large τ even at the larger values of θ , where the variance error is comparatively small. The question arises as to why a defensive strategy would be employed in the presence of large bias and relatively small variance error. The population model surface was fairly complex as suggested by equation (26) and a maximum point was located at $\delta_k = -1, 1, -1, -1, \dots$, but that point was not one of the observation points of the fractional factorial experiments. The irregular shape of

the population model could cause relatively large magnitude coefficients of the higher order terms of the fitted models to be retained, which would bring predicted values close to the population values for those cases where predictions were being made at points of the design where "observations" had been recorded. For the points of the space corresponding to fractions of the experiment not performed, the predictions could contain very large errors due to such high order terms. For these predictions, the errors are reduced by the deletion of such terms and this is why the optimal strategies of table XIII were highly defensive (used small values of α_p and α_f) for $h > 0$ when both θ was large and τ was ≥ 1.0 .

Aside from questions arising from $\tau > 1$ two questions that might be asked are: "Under what conditions are some of the $(m_p, \alpha_p, \alpha_f)$ strategies optimal?" and "Is there some particular strategy $(m_p, \alpha_p, \alpha_f)$ that is widely useful?" An attempt to exhibit an answer to these questions is given in table XIV. It lists α_p and α_f as row and column headings and lists the parameter values g and θ as tabular entries. Results for the purely first degree population model ($\tau = 0$) are shown by table XIV(a) and results in the presence of a small amount of surface curvature ($\tau = 1$) are shown by table XIV(b). These results show that for $\tau = 0$ and θ small ($\theta \leq 4.0$) the strategy $(m_p, \alpha_p, \alpha_f) = (1, 0.75, 0.10)$ is a widely useful strategy, but that for $\tau = 0$ and large θ , the optimal strategy most often used $\alpha_p = \alpha_f$, with decreasing values of α_p and α_f for increasing values of θ . In comparison with these results for $\tau = 0$, table XIV(b) showed that relatively larger values of α_p were optimal for $\tau = 1$. (Large values of α_p tend to inhibit the pooling of mean squares into the denominator of the test statistic, which was apparently beneficial when the bias risk was increased by going from $\tau = 0.0$ to $\tau = 1.0$).

The strategy $(m_p, \alpha_p, \alpha_f) = (1, 0.75, 0.10)$ was suggested as a widely useful strategy, particularly for those values of θ ($1.0 \leq \theta \leq 4.0$) where the relative influence of the variance error would be moderate and for the value of τ ($\tau = 0.0$) where the bias error would be absent. Other values of θ and τ (table XIV) had resulted in other strategies being optimal. These results raise the question of: "Just how fast does the mean square prediction error and its variance change when the values of α_p and α_f are other than optimal?"

Some parameter values of g and θ requiring relatively extreme values of m_p , α_p , and α_f are designated by the footnoted values in table XIV. How fast the mean square prediction error and its variance changes when the value of α_p and α_f are other than optimal for these footnoted cases of g and θ are exhibited by table XV. (Some values of g and θ having less extreme values of α_p and α_f as optimal strategies are also exhibited in table XV.) The cases exhibited in table XV are limited to the low bias cases of $\tau \leq 1.0$. The values of \bar{e}_{\max}^2 that identify optimal values of α_p and α_f are indicated by solid line rectangles. For comparison, the values of \bar{e}_{\max}^2 associated with the widely optimal strategy of $(m_p, \alpha_p, \alpha_f) = (1, 0.75, 0.10)$ are

identified by dashed line rectangles. One conclusion that is evidenced by table XV, is that the values of \bar{e}_{\max}^2 and $V(\bar{e}^2)_{\max}$ always increase very slowly from the optimal values of α_p and α_f for the more aggressive strategies (larger values of α_p and α_f). On the other hand (see particularly tables XV(c) and (g)), \bar{e}_{\max}^2 and $V(\bar{e}^2)_{\max}$ increase quite rapidly for strategies that are too defensive (values of α_p and α_f that are smaller than optimal).

These observations suggest that the use of a moderate strategy when an aggressive strategy is optimal is not too serious - the differences are mainly functions of variance error. On the other hand the use of a defensive strategy when it is not optimal (as in some cases of large bias) can be disastrous.

The overall conclusions from the results exhibited by tables XIII, XIV, and XV seem to be that chain pooling with 16 orthogonal estimates of regression coefficients with no estimate of pure error should be done with the strategy of $(m_p, \alpha_p, \alpha_f) = (1, 0.75, 0.10)$. Such a strategy is a moderate strategy tending to retain many of the terms, if many terms are needed. If, with such a strategy, many of the terms are rejected, the true situation may be concluded to be one where the underlying population response surface is relatively simple. In such a situation, the fitted equation may be used (1) for data smoothing (for computing predicted values at points of the space that gave the observed values), (2) for interpolation (for computing predicted values at points interior to the space spanned by the experiment, and (3) for extrapolation (for computing predicted values at points exterior to the space of the experiment). Of course - as is well known for extrapolations - even if the bias error is small, the variance error rises very rapidly with the distance of the predicted points from the space spanned by the observed points).

If on the other hand, the strategy $(m_p, \alpha_p, \alpha_f) = (1, 0.75, 0.10)$ leads to the retention of a fairly complex model equation (one with several terms of higher than first degree) then the use of a more defensive strategy might result in an equation with less bias when used for interpolation or extrapolation, but the prediction errors would be large in any case (as shown by table XIII for $g > 4$ and $\tau > 1$). In these circumstances, the reasonable conclusion would seem to be that the reduced model obtained with $(m_p, \alpha_p, \alpha_f) = (1, 0.75, 0.10)$ is satisfactory for data smoothing at the treatment points, but that more experimenting must be done (such as the performance of the composite experiments of Box and Hunter (1957)) if any interpolation is to be done.

CONCLUDING REMARKS

Model fitting procedures were investigated for the types and sizes of experiments that are appropriate to empirical materials optimization studies. The two-level, fractional-factorial, fixed effects experiments are often appropriate to such purposes, and they provide the highly desirable orthogonal estimates of the model parameters. Chain pooling is

an appropriate method of parameter estimation and selection for such experiments.

For the cases of nonorthogonal experiments, the subject of principal components regression was examined and was found to provide a suitable procedure for making the results of such experiments amenable to the methods of chain pooling.

In empirical optimum seeking (as in materials optimization) the criterion for model parameter estimation and selection should be the minimization of prediction error. In the Monte Carlo study performed to find minimum prediction error strategies when chain pooling methods are used for model selection, simulations were performed using population models intended to represent response surface applications. The results led to the following recommendations:

1. Use the strategy $(m_p, \alpha_p, \alpha_f) = (1, 0.75, 0.10)$.
2. If many terms are deleted - if only a few terms of higher than first degree are retained - use the reduced model for data smoothing, interpolation, and limited extrapolation.
3. If only a few terms are deleted - if more than a few terms of higher than first degree are retained - use the reduced model for data smoothing only. Additional observations are needed for predictions at points other than those already observed.

APPENDIX A

SYMBOLS

FORTRAN	Mathematical symbol	
I	i	Subscript on treatments
IT	t	Number of treatments
J	j	Subscript on ordered mean squares
K	k	Subscript on independent variables
KG	g	Number of independent variables
L	ℓ	Subscript on parameter θ
LTH	ℓ_θ	Number of values of θ
M	m	Subscript on parameter τ
MTAU	m_τ	Number of values of τ
N	n	Subscript on experiments
NE	n_e	Number of experiments
MP	m_p	Number of mean squares pooled initially
LGMH	$g-h$	Experiment contains 2^{g-h} treatments
PHI	ϕ_k	Exponential parameters, $k = 1, \dots, g$
PSI	ψ_k	First degree parameters, $k = 1, \dots, g$
THETA	θ_ℓ	Scale parameters, $\ell = 1, \dots, \ell_\theta$
TAU	τ_m	Exponential parameter, $m = 1, \dots, m_\tau$
YMU	$\mu_{i\ell m}$	Population mean for i^{th} treatment, ℓ^{th} value of θ_ℓ and m^{th} value of τ_m
	x_{ik}	Value of k^{th} independent variable at i^{th} treatment. For two level fractional factorial experiments; at its upper level, $x_{ik} = 1$; at its lower level, $x_{ik} = -1$.
	α_p	Nominal significance level of preliminary test

FORTRAN name	Mathematical symbol	
	α_f	Nominal significance level of final test
KP		Column number of U_j table associated with specified α_p
KF		Column number of U_j table associated with specified α_f

APPENDIX B

PROGRAM POOLLES

1.- DECLARATIONS, CONSTANTS, POPULATION MEANS, AND STRATEGY

```

DIMENSION REMARK(14), ALPHA(11), TB(64,10), PHI(8),
1PSI(8), THETA(6), TAU(6), YMU(128,6,6), RN(64), IND(64), Z(64),
ZERSQ(128,6,6), ERSQSQ(128,6,6), ERSQMX(6,6), AVERSQ(6,6),
3VESQMX(6,6), AVVESQ(6,6), DEL(8)
COMMON KK, YOBS(256), B(256)
DATA (ALPHA(I), I=1,11)/0.001,0.002,0.005,0.01,0.025,0.05,0.10,
10.25,0.50,0.75,1.0/
DATA ((TB(I,J), J=1,10), I=1,24)/0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,
1,0.0,2.0000,1.99999,1.99997,1.99986,1.99917,1.99687,1.9877,1.923,1
2.706,1.382,2.9976,2.9960,2.9904,2.9809,2.951,2.904,2.806,2.527,2.0
386,1.688,3.976,3.962,3.925,3.870,3.760,3.625,3.412,2.949,2.395,1.9
461,4.887,4.845,4.758,4.65,4.44,4.21,3.89,3.287,2.658,2.184,5.74,5.
563,5.46,5.31,4.99,4.68,4.28,3.57,2.893,2.371,6.51,6.33,6.11,5.87,5
6.46,5.09,4.61,3.83,3.11,2.54,7.20,6.96,6.65,6.35,5.88,5.44,4.91,4.
706,3.29,2.69,7.81,7.52,7.10,6.78,6.26,5.75,5.17,4.27,3.45,2.82,8.3
84,8.01,7.53,7.17,6.59,6.03,5.41,4.45,3.60,2.95,8.82,8.44,7.95,7.53
9,6.89,6.28,5.61,4.62,3.74,3.07,9.26,8.84,8.33,7.87,7.13,6.50,5.81,
A4.77,3.87,3.17,9.67,9.21,8.68,8.16,7.37,6.71,5.99,4.92,3.99,3.27,1
B0.05,9.55,8.95,8.42,7.59,6.91,6.15,5.05,4.10,3.37,10.40,9.86,9.20,
C8.66,7.79,7.07,6.30,5.17,4.20,3.46,10.72,10.14,9.43,8.83,7.96,7.23
D,6.44,5.29,4.30,3.55,11.01,10.40,9.64,9.00,8.12,7.38,6.57,5.40,4.3
E9,3.63,11.28,10.64,9.84,9.17,8.28,7.52,6.69,5.50,4.48,3.70,11.53,1
F0.86,10.03,9.34,8.43,7.65,6.81,5.60,4.56,3.77,11.76,11.07,10.22,9.
G51,8.58,7.78,6.92,5.69,4.64,3.84,11.98,11.28,10.40,9.67,8.72,7.90,
H7.03,5.78,4.71,3.90,12.19,11.48,10.58,9.83,8.86,8.02,7.13,5.87,4.7
I8,3.96,12.39,11.68,10.76,9.99,8.99,8.13,7.23,5.95,4.85,4.02,12.58,
J11.87,10.93,10.14,9.12,8.24,7.33,6.03,4.92,4.08/
DATA ((TB(I,J), J=1,10), I=25,47)/12.76,12.05,11.10,10.29,9.23,8.34,
17.42,6.11,4.98,4.14,12.93,12.22,11.26,10.43,9.34,8.44,7.51,6.18,5.
204,4.19,13.09,12.38,11.41,10.56,9.44,8.54,7.60,6.25,5.10,4.24,13.2
34,12.53,11.55,10.68,9.54,8.63,7.68,6.32,5.16,4.30,13.39,12.68,11.6
48,10.78,9.64,8.72,7.76,6.38,5.22,4.35,13.53,12.82,11.80,10.88,9.74
5,8.81,7.83,6.44,5.28,4.40,13.67,12.96,11.91,10.98,9.83,8.89,7.90,6
6.50,5.33,4.45,13.80,13.09,12.01,11.07,9.91,8.97,7.97,6.56,5.38,4.5
70,13.93,13.21,12.10,11.16,9.99,9.04,8.04,6.62,5.43,4.54,14.05,13.3
82,12.19,11.25,10.07,9.11,8.11,6.68,5.48,4.58,14.17,13.43,12.27,11.
934,10.15,9.18,8.17,6.74,5.53,4.62,14.29,13.53,12.35,11.43,10.22,9.
A25,8.23,6.80,5.58,4.66,14.41,13.63,12.43,11.51,10.29,9.31,8.29,6.8
B5,5.63,4.70,14.53,13.73,12.51,11.59,10.36,9.37,8.35,6.90,5.67,4.74
C,14.64,13.82,12.59,11.67,10.43,9.43,8.41,6.95,5.71,4.78,14.75,13.9
D1,12.67,11.75,10.50,9.49,8.46,6.99,5.75,4.82,14.85,14.00,12.75,11.
E83,10.57,9.55,8.51,7.03,5.79,4.86,14.95,14.09,12.83,11.90,10.64,9.
F61,8.56,7.07,5.83,4.90,15.05,14.17,12.90,11.97,10.70,9.67,8.61,7.1
G1,5.87,4.94,15.15,14.25,12.97,12.04,10.76,9.72,8.66,7.15,5.91,4.98
H,15.24,14.33,13.05,12.11,10.82,9.77,8.71,7.19,5.95,5.01,15.33,14.4
I0,13.12,12.18,10.88,9.82,8.76,7.23,5.99,5.04,15.42,14.47,13.19,12.
J25,10.94,9.87,8.81,7.27,6.03,5.07/
DATA((TB(I,J), J=1,10), I=48,64)/15.50,14.54,13.26,12.32,11.00,9.92,

```

```

18.85,7.31,6.07,5.10,15.58,14.60,13.32,12.38,11.06,9.97,8.89,7.35,6
2.11,5.13,15.66,14.66,13.38,12.44,11.11,10.02,8.93,7.39,6.14,5.16,1
35.73,14.72,13.44,12.50,11.16,10.07,8.97,7.43,6.17,5.19,15.80,14.79
4,13.50,12.56,11.21,10.12,9.01,7.47,6.20,5.22,15.87,14.85,13.56,12.
562,11.26,10.17,9.05,7.51,6.23,5.25,15.93,14.91,13.62,12.68,11.31,1
60.21,9.09,7.55,6.26,5.28,15.99,14.97,13.67,12.73,11.36,10.25,9.13,
77.59,6.29,5.31,16.05,15.03,13.72,12.78,11.40,10.29,9.17,7.63,6.32,
85.34,16.11,15.10,13.77,12.83,11.44,10.33,9.21,7.67,6.35,5.37,16.17
9,15.16,13.82,12.88,11.48,10.37,9.25,7.70,6.38,5.40,16.23,15.22,13.
A87,12.93,11.52,10.41,9.29,7.73,6.41,5.43,16.29,15.28,13.92,12.97,1
61.56,10.45,9.33,7.76,6.44,5.46,16.34,15.34,13.97,13.01,11.60,10.49
C,9.37,7.79,6.47,5.48,16.39,15.40,14.02,13.05,11.64,10.53,9.41,7.82
D,6.50,5.50,16.44,15.46,14.06,13.09,11.67,10.57,9.45,7.85,6.53,5.52
E,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0/

```

C

```

READ(5,800) (REMARK(I),I=1,14)
WRITE(6,801) (REMARK(I),I=1,14)
READ(5,824) LGMH, NE
IT= 2**LGMH
ITM1= IT-1
ITM2 = IT-2
FIT= IT
FNE= NE
IF (NE.EQ.0) FNE = 1.0
FNEM1 = FNE - 1.0
READ (5,825) KG, (PSI(K),K=1,KG)
WRITE (6,842) (PSI(K), K=1,KG)
READ (5,825) KG, (PHI(K),K=1,KG)
WRITE (6,843) (PHI(K), K=1,KG)
READ (5,825) KG, (DEL(K), K=1,KG)
WRITE (6,844) (DEL(K), K=1,KG)
READ(5,825) LTH, (THETA(L), L=1,LTH)
READ(5,825) MTAU, (TAU(M), M=1,MTAU)
IP= 2**KG
FIP = IP
IPPI= IP + 1
FFNIP = FNE*FIP
FEM1IP = FNEM1 * FIP

```

C

```

DO 49 M=1,MTAU
DO 48 L=1,LTH
DO 20 I=1,IP
A= 1.0
C = 0.0
DO 15 K=1,KG
KM1= K-1
J= 2**KM1
AX = AND((I-1),J)
XI = 2*IARS(KM1,AX)-1
DEBUG I, K, XI
A= A+PSI(K)*XI
C = C + PHI(K)*(XI-DEL(K))*2
15 CONTINUE
YMU(I,L,M) = THETA(L)*(A+TAU(M)*EXP(-C))
DEBUG I, L, M, YMU(I,L,M)
20 CONTINUE

```

48 CONTINUE
49 CONTINUE

C

50 READ (5,824) MP,KP,KF
MPPI = MP+1
CALL SAND (XS)
DO 99 M=1,MTAU
DO 98 L=1,LTH
DO 97 I=1,IP
ERSQ(I,L,M)= 0.0
ERSQSQ(I,L,M)= 0.0
97 CONTINUE
98 CONTINUE
99 CONTINUE

C

C 2.- SIMULATIONS AND MODEL FITTING

C

DO 699 N=1,NE
IF (NE.EQ.0) GO TO 200
DO 213 I=1,IT
CALL RAND(RN(I))
DEBUG I, RN(I)
213 CONTINUE
DO 215 I=1,IT,2
E= SQRT(-2.0*ALOG(RN(I)))
D= 6.2831853*RN(I+1)
FN(I)= E*COS(D)
DEBUG RN(I)
FN(I+1)= E*SIN(D)
DEBUG RN(I+1)
215 CONTINUE
GO TO 201
200 DO 209 I=1,IT
FN(I) = 0.0
209 CONTINUE
201 DO 690 M=1,MTAU
DO 680 L=1,LTH
KK= LGMH
IF(KG .EQ. 4) GO TO 204
IF(KG .EQ. 5) GO TO 205
IF(KG .EQ. 6) GO TO 206
IF(KG .EQ. 7) GO TO 207
IF(KG .EQ. 8) GO TO 208

C

204 DO 214 I=1,IT
Y OBS(I)= YMU(I,L,M) +RN(I)
DEBUG Y OBS (I)
214 CONTINUE
CALL YATES
GO TO 300

C

205 Y OBS(1) = YMU(1,L,M)+RN(1)
Y OBS(2) = YMU(2,L,M)+RN(2)
Y OBS(3) = YMU(19,L,M)+RN(3)
Y OBS(4) = YMU(20,L,M)+RN(4)
Y OBS(5) = YMU(21,L,M)+RN(5)

```

YOB( 6) = YMU(22,L,M)+RN( 6)
YOB( 7) = YMU( 7,L,M)+RN( 7)
YOB( 8) = YMU( 8,L,M)+RN( 8)
YOB( 9) = YMU(25,L,M)+RN( 9)
YOB(10) = YMU(26,L,M)+RN(10)
YOB(11) = YMU(11,L,M)+RN(11)
YOB(12) = YMU(12,L,M)+RN(12)
YOB(13) = YMU(13,L,M)+RN(13)
YOB(14) = YMU(14,L,M)+RN(14)
YOB(15) = YMU(31,L,M)+RN(15)
YOB(16) = YMU(32,L,M)+RN(16)

```

```
DO 250 I=1,IT
```

```
  DEBUG YOB(I)
```

```
250 CONTINUE
```

```
  CALL YATES
```

```
  GO TO 300
```

C

```

206 YOB( 1) = YMU( 1,L,M)+RN( 1)
YOB( 2) = YMU(34,L,M)+RN( 2)
YOB( 3) = YMU(19,L,M)+RN( 3)
YOB( 4) = YMU(52,L,M)+RN( 4)
YOB( 5) = YMU(53,L,M)+RN( 5)
YOB( 6) = YMU(22,L,M)+RN( 6)
YOB( 7) = YMU(39,L,M)+RN( 7)
YOB( 8) = YMU( 8,L,M)+RN( 8)
YOB( 9) = YMU(57,L,M)+RN( 9)
YOB(10) = YMU(26,L,M)+RN(10)
YOB(11) = YMU(43,L,M)+RN(11)
YOB(12) = YMU(12,L,M)+RN(12)
YOB(13) = YMU(13,L,M)+RN(13)
YOB(14) = YMU(46,L,M)+RN(14)
YOB(15) = YMU(31,L,M)+RN(15)
YOB(16) = YMU(64,L,M)+RN(16)

```

```
DO 260 I=1,IT
```

```
  DEBUG YOB(I)
```

```
260 CONTINUE
```

```
  CALL YATES
```

```
  GO TO 300
```

C

```

207 YOB( 1) = YMU( 1,L,M)+RN( 1)
YOB( 2) = YMU( 98,L,M)+RN( 2)
YOB( 3) = YMU( 83,L,M)+RN( 3)
YOB( 4) = YMU( 52,L,M)+RN( 4)
YOB( 5) = YMU( 53,L,M)+RN( 5)
YOB( 6) = YMU( 86,L,M)+RN( 6)
YOB( 7) = YMU(103,L,M)+RN( 7)
YOB( 8) = YMU( 8,L,M)+RN( 8)
YOB( 9) = YMU(121,L,M)+RN( 9)
YOB(10) = YMU( 26,L,M)+RN(10)
YOB(11) = YMU( 43,L,M)+RN(11)
YOB(12) = YMU( 76,L,M)+RN(12)
YOB(13) = YMU( 77,L,M)+RN(13)
YOB(14) = YMU( 46,L,M)+RN(14)
YOB(15) = YMU( 31,L,M)+RN(15)
YOB(16) = YMU(128,L,M)+RN(16)
DO 270 I=1,IT

```



```

      DEBUG YOBS(I)
270  CONTINUE
      CALL YATES
      GO TO 300

```

C

```

208  YOBS( 1) = YMU( 1,L,M)+RN( 1)
      YOBS( 2) = YMU(226,L,M)+RN( 2)
      YOBS( 3) = YMU(211,L,M)+RN( 3)
      YOBS( 4) = YMU( 52,L,M)+RN( 4)
      YOBS( 5) = YMU(181,L,M)+RN( 5)
      YOBS( 6) = YMU( 86,L,M)+RN( 6)
      YOBS( 7) = YMU(103,L,M)+RN( 7)
      YOBS( 8) = YMU(136,L,M)+RN( 8)
      YOBS( 9) = YMU(121,L,M)+RN( 9)
      YOBS(10) = YMU(154,L,M)+RN(10)
      YOBS(11) = YMU(171,L,M)+RN(11)
      YOBS(12) = YMU( 76,L,M)+RN(12)
      YOBS(13) = YMU(205,L,M)+RN(13)
      YOBS(14) = YMU( 46,L,M)+RN(14)
      YOBS(15) = YMU( 31,L,M)+RN(15)
      YOBS(16) = YMU(256,L,M)+RN(16)
      DO 280 I=1,IT
      DEBUG YOBS(I)
280  CONTINUE
      CALL YATES
      GO TO 300

```

C

C

C

3.- CONSTRUCTION AND ORDERING OF MEAN SQUARES

```

300  DO 309 I=1,IT
      IND(I)= I
      Z(I)= B(I+1)*B(I+1)/FIT
      B(I)= B(I) / FIT
      DEBUG B(I)
309  CONTINUE
      IF (MP.EQ.0) GO TO 500

```

C

```

      DO 313 J=1,ITM2
      TEST= Z(ITM1)
      IN= ITM1
      DO 312 NA=J,ITM2
      IF(TEST-Z(NA)) 312,312,311
311  TEST = Z(NA)
      IN= NA
312  CONTINUE
      ITEM= IND(IN)
      TEM= Z(IN)
      IND(IN)= IND(J)
      Z(IN)= Z(J)
      IND(J)= ITEM
      Z(J)= TEM
      DEBUG IND(J),Z(J),B(ITEM)
313  CONTINUE
      DEBUG IND(15),Z(15),B(15)

```

C

C

4.- DELETION OF TERMS

C

```

JN= MPPI
TEM= 0.0
DO 415 J=1,MP
INDX = IND(J)+1
B(INDX)= 0.0
TEM= TEM + Z(J)
DEBUG JN, INDX,B(INDX), TEM
415 CONTINUE
IF (KF.EQ.11) GO TO 500
DO 419 J=MPPI,ITM1
FJN = JN
TEST = FJN*Z(J)/(TEM+Z(J))
IF (KP.EQ.11) GO TO 417
IF (TEST-TB(JN,KP)) 416,416,417
416 TEM = TEM + Z(J)
INDX = IND(J)+1
B(INDX) = 0.0
JN = JN+1
DEBUG JN,INDX,B(INDX),TEM
GO TO 419
417 IF(TEST - TB(JN,KP)) 418,418,420
418 INDX = IND(J)+1
B(INDX) = 0.0
DEBUG INDX, B(INDX),TEM
419 CONTINUE
JETA = ITM1
GO TO 500
420 JETA = J-1
DO 421 J=JETA,IT
INDX = IND(J)+1
DEBUG INDX, B(INDX)
421 CONTINUE

```

C

C

C

5.- PREDICTIONS

```

500 KK = KG
DEBUG JETA
IF (KG .EQ. 4) GO TO 540
IF (KG .EQ. 5) GO TO 550
IF (KG .EQ. 6) GO TO 560
IF (KG .EQ. 7) GO TO 570
IF (KG .EQ. 8) GO TO 580
540 DO 546 I=1,IP
IPP1MI = IPP1-I
YOBS(I) = B(IPP1MI)
DEBUG YOBS(I)
546 CONTINUE
CALL YATES
GO TO 600
550 DO 551 I=1,20
551 YOBS(I) = 0.0
YOBS(13) = B(14)
YOBS(14) = B(13)
YOBS(15) = B(16)
YOBS(16) = B(15)

```

```

      DO 552 I=21,32
      IPP1MI = 33-I
552  YOBS(I) = B(IPP1MI)
      DO 557 I=1,IP
      DEBUG YOBS(I)
557  CONTINUE
      CALL YATES
      GO TO 600

```

C

```

560 DO 561 I=1,52
561  YOBS(I) = 0.0
      YOBS(31) = B(13)
      YOBS(32) = B(14)
      YOBS(47) = B(16)
      YOBS(48) = B(15)
      DO 562 I=53,64
      IPP1MI = 65-I
562  YOBS(I) = B(IPP1MI)
      DO 567 I=1,IP
      DEBUG YOBS(I)
567  CONTINUE
      CALL YATES
      GO TO 600

```

C

```

570 DO 571 I=1,117
571  YOBS(I) = 0.0
      YOBS(64) = B(12)
      YOBS(95) = B(13)
      YOBS(96) = B(14)
      YOBS(111) = B(16)
      YOBS(112) = B(15)
      DO 572 I=118,128
      IPP1MI = 129-I
572  YOBS(I) = B(IPP1MI)
      DO 577 I=1,IP
      DEBUG YOBS(I)
577  CONTINUE
      CALL YATES
      GO TO 600

```

C

```

580 DO 581 I=1,245
581  YOBS(I) = 0.0
      YOBS(128) = B(8)
      YOBS(192) = B(12)
      YOBS(223) = B(13)
      YOBS(224) = B(14)
      YOBS(239) = B(16)
      YOBS(240) = B(15)
      DO 582 I=246,256
      IPP1MI = 257-I
582  YOBS(I) = B(IPP1MI)
      YOBS(249) = 0.0
      DO 587 I=1,IP
      DEBUG YOBS(I)
587  CONTINUE
      CALL YATES

```

GO TO 600

C
C
C

6.- ACCUMULATION OF ERRORS

```
600 DO 639 I=1,IP
    IPP1MI = IPP1-I
    DEBUG B(IPP1MI)
    TEM= (B(IPP1MI) - YMU(I,L,M))**2
    ERSQ(I,L,M) = ERSQ(I,L,M) + TEM
    ERSQSQ(I,L,M)= ERSQSQ(I,L,M) + TEM**2
    DEBUG M,L,I,ERSQ(I,L,M)
609 CONTINUE
680 CONTINUE
690 CONTINUE
    IF (NE .EQ. 0) GO TO 700
699 CONTINUE
```

C
C
C
C

7.- DETERMINATION OF MAXIMUM AND MEAN SQUARED ERRORS

```
700 DO 790 M=1,MTAU
    DO 780 L=1,LTH
        C = 0.0
        D = 0.0
        E = 0.0
        F = 0.0
        DO 750 I=1,IP
            C = AMAX1(C,ERSQ(I,L,M))
            D = D + ERSQ(I,L,M)
            TEM= ERSQSQ(I,L,M)- ((ERSQ(I,L,M))**2)/FNE
            E = AMAX1(E,TEM)
            F = F + TEM
        750 CONTINUE
        ERSQMX(L,M) = C/FNE
        AVERSQ(L,M) = D/FENBIP
        IF (NE .EQ. 0) GO TO 780
        VESQMX(L,M)= E/FNEM1
        AVVESQ(L,M)= F/FEM1IP
        DEBUG ERSQMX(L,M),AVERSQ(L,M),VESQMX(L,M),AVVESQ(L,M)
    780 CONTINUE
    790 CONTINUE
```

C
C
C

8.- OUTPUT

```
WRITE (6,830) IT,NE,KG,MP,ALPHA(KP),ALPHA(KF)
WRITE (6,832) (TAU(M),M=1,MTAU)
WRITE (6,833)
WRITE (6,831)
WRITE (6,834) (THETA(L),(ERSQMX(L,M),M=1,MTAU),L=1,LTH)
WRITE (6,835)
WRITE (6,834) (THETA(L),(AVERSQ(L,M),M=1,MTAU),L=1,LTH)
WRITE (6,836)
WRITE (6,834) (THETA(L),(VESQMX(L,M),M=1,MTAU),L=1,LTH)
WRITE(6,837)
WRITE(6,834) (THETA(L),(AVVESQ(L,M),M=1,MTAU),L=1,LTH)
GO TO 50
```

C

```
800 FORMAT (13A6,A2)
801 FORMAT (1H1,10X,13A6,A2//)
824 FORMAT (3I5)
825 FORMAT (I5,8F5.0)
830 FORMAT (1H1,3X,4HIT =I5,5X,4HNE =I5,5X,4HKG =I5,6X,4HMP =I5,5X,
  18HALPHAP =F6.3,5X,8HALPHAF =F6.3//)
831 FORMAT (1H0,20X,6HERSQMX//)
832 FORMAT (1H0,4HTAU=6F13.2//)
833 FORMAT (1H0,5HTHETA//)
834 FORMAT (F6.2,6E14.4)
835 FORMAT (1H0,20X,6HAVERSQ//)
836 FORMAT (1H0,20X,6HVESQMX//)
837 FORMAT (1H0,20X,6HAVVESQ//)
842 FORMAT (1H0,4HPSI=8E14.4)
843 FORMAT (1H0,4HPHI=8E14.4)
844 FORMAT (1H0,4HDEL=8E14.4)
  END
```

SUBROUTINE YATES

C
C 9.- YATES METHOD SUBROUTINE
C

COMMON KK,Y(256),B(256)

II = 2**KK

IIOB2 = II/2

KKM1 = KK-1

DO 908 K=1,KKM1

DO 906 I=1,II,2

IP1D2 = (I+1)/2

B(IP1D2) = Y(I+1)+Y(I)

LL = IP1D2+IIOB2

906 B(LL) = Y(I+1)-Y(I)

DO 907 I=1,II

907 Y(I) = B(I)

908 CONTINUE

DO 909 I=1,II,2

IP1D2 = (I+1)/2

B(IP1D2) = Y(I+1)+Y(I)

LL = IP1D2+IIOB2

B(LL) = Y(I+1)-Y(I)

909 CONTINUE

RETURN

END

APPENDIX C

ILLUSTRATIVE OUTPUT - POOLLES

TAU=	0.	0.50	1.00	1.50	2.00	2.50
THETA						
ERSQMX						
0.50	0.1673E+01	0.1673E+01	0.1673E+01	0.1673E+01	0.1673E+01	0.1673E+01
0.25	0.1673E+01	0.1673E+01	0.1673E+01	0.1673E+01	0.1673E+01	0.1673E+01
0.13	0.1673E+01	0.1673E+01	0.1673E+01	0.1673E+01	0.1673E+01	0.1673E+01
AVERSQ						
0.50	0.8256E+00	0.8256E+00	0.8256E+00	0.8256E+00	0.8256E+00	0.8256E+00
0.25	0.8256E+00	0.8256E+00	0.8256E+00	0.8256E+00	0.8256E+00	0.8256E+00
0.13	0.8256E+00	0.8256E+00	0.8256E+00	0.8256E+00	0.8256E+00	0.8256E+00
VESQMX						
0.50	0.3585E+01	0.3585E+01	0.3585E+01	0.3585E+01	0.3585E+01	0.3585E+01
0.25	0.3585E+01	0.3585E+01	0.3585E+01	0.3585E+01	0.3585E+01	0.3585E+01
0.13	0.3585E+01	0.3585E+01	0.3585E+01	0.3585E+01	0.3585E+01	0.3585E+01
AVVESQ						
0.50	0.1121E+01	0.1121E+01	0.1121E+01	0.1121E+01	0.1121E+01	0.1121E+01
0.25	0.1121E+01	0.1121E+01	0.1121E+01	0.1121E+01	0.1121E+01	0.1121E+01
0.13	0.1121E+01	0.1121E+01	0.1121E+01	0.1121E+01	0.1121E+01	0.1121E+01

APPENDIX D

PROGRAM MODEL

1.- DECLARATIONS AND INPUT

```

DIMENSION REMARK(14), PSI(8), PHI(8), THETA(6), TAU(6), ALPHA(11),
IYMU(256,6,6), BIAS(256,6,6), BILM(16,6,6), DEL(8)
COMMON KK, YORS(256), B(256)
DATA (ALPHA(I), I=1,11)/0.001,0.002,0.005,0.01,0.025,0.05,0.10,
10.25,0.50,0.75,1.0/
READ(5,800) (REMARK(I), I=1,14)
WRITE(6,801) (REMARK(I), I=1,14)
READ(5,824) LGMH, NE
IT= 2**LGMH
FIT= IT
READ (5,825) KG, (PSI(K), K=1,KG)
WRITE (6,842) (PSI(K), K=1,KG)
READ (5,825) KG, (PHI(K), K=1,KG)
WRITE (6,843) (PHI(K), K=1,KG)
READ (5,825) KG, (DEL(K), K=1,KG)
WRITE (6,844) (DEL(K), K=1,KG)
READ(5,825) LTH, (THETA(L), L=1,LTH)
READ(5,825) MTAU, (TAU(M), M=1,MTAU)
IP= 2**KG
FIP = IP
IPP1= IP + 1

```

2.- MODEL AND FITTING

```

DO 49 M=1,MTAU
DO 48 L=1,LTH
DO 20 I=1,IP
A= 1.0
C = 0.0
DO 15 K=1,KG
KM1= K-1
J= 2**KM1
AX = AND((I-1),J)
XI = 2*IARS(KM1,AX)-1
A= A+PSI(K)*XI
C = C + PHI(K)*(XI-DEL(K))*2
15 CONTINUE
YMU(I,L,M) = THETA(L)*(A+TAU(M)*EXP(-C))
BIAS(I,L,M) = 0.0
20 CONTINUE
48 CONTINUE
49 CONTINUE
50 READ (5,824) MP,KP,KF

```

3.- PREDICTIONS

```

DO 690 M=1,MTAU
DO 680 L=1,LTH

```



```

KK= LGMH
IF(KG .EQ. 4) GO TO 204
IF(KG .EQ. 5) GO TO 205
IF(KG .EQ. 6) GO TO 206
IF(KG .EQ. 7) GO TO 207
IF(KG .EQ. 8) GO TO 208

```

C

```

204 DO 214 I=1,IT
  YORS(I)= YMU(I,L,M)
214 CONTINUE
  CALL YATES
  DO 349 I=1,IT
    B(I) = B(I)/FIT
    BILP(I,L,M) = B(I)
349 CONTINUE

```

C

```

  KK = KG
  DO 546 I=1,IP
    IPP1M1 = IPP1-I
    YORS(I) = B(IPP1M1)
546 CONTINUE
  CALL YATES
  GO TO 600

```

C

```

205 YORS( 1) = YMU( 1,L,M)
  YORS( 2) = YMU( 2,L,M)
  YORS( 3) = YMU(19,L,M)
  YORS( 4) = YMU(20,L,M)
  YORS( 5) = YMU(21,L,M)
  YORS( 6) = YMU(22,L,M)
  YORS( 7) = YMU( 7,L,M)
  YORS( 8) = YMU( 8,L,M)
  YORS( 9) = YMU(25,L,M)
  YORS(10) = YMU(26,L,M)
  YORS(11) = YMU(11,L,M)
  YORS(12) = YMU(12,L,M)
  YORS(13) = YMU(13,L,M)
  YORS(14) = YMU(14,L,M)
  YORS(15) = YMU(31,L,M)
  YORS(16) = YMU(32,L,M)
  CALL YATES
  DO 359 I=1,IT
    B(I) = B(I)/FIT
    BILP(I,L,M) = B(I)
359 CONTINUE
  KK = KG
  DO 551 I=1,20
551 YORS(I) = 0.0
  YORS(13) = B(14)
  YORS(14) = B(13)
  YORS(15) = B(16)
  YORS(16) = B(15)
  DO 552 I=21,32
    IPP1M1 = 33-I
552 YORS(I) = B(IPP1M1)
  CALL YATES

```

GO TO 600

C

```

206 YORS( 1) = YMU( 1,L,M)
YORS( 2) = YMU(34,L,M)
YORS( 3) = YMU(19,L,M)
YORS( 4) = YMU(52,L,M)
YORS( 5) = YMU(53,L,M)
YORS( 6) = YMU(22,L,M)
YORS( 7) = YMU(39,L,M)
YORS( 8) = YMU( 8,L,M)
YORS( 9) = YMU(57,L,M)
YORS(10) = YMU(26,L,M)
YORS(11) = YMU(43,L,M)
YORS(12) = YMU(12,L,M)
YORS(13) = YMU(13,L,M)
YORS(14) = YMU(46,L,M)
YORS(15) = YMU(31,L,M)
YORS(16) = YMU(64,L,M)
CALL YATES
DO 369 I=1,IT
B(I) = B(I)/FIT
BILP(I,L,M)= B(I)
369 CONTINUE
KK = KG
DO 561 I=1,52
561 YORS(I) = 0.0
YORS(31) = B(13)
YORS(32) = B(14)
YORS(47) = B(16)
YORS(48) = B(15)
DO 562 I=53,64
IPFIM I = 65-I
562 YORS(I) = B(IPFIM I)
CALL YATES
GO TO 600

```

C

```

207 YORS( 1) = YMU( 1,L,M)
YORS( 2) = YMU( 98,L,M)
YORS( 3) = YMU( 83,L,M)
YORS( 4) = YMU( 52,L,M)
YORS( 5) = YMU( 53,L,M)
YORS( 6) = YMU( 66,L,M)
YORS( 7) = YMU(103,L,M)
YORS( 8) = YMU( 8,L,M)
YORS( 9) = YMU(121,L,M)
YORS(10) = YMU( 26,L,M)
YORS(11) = YMU( 43,L,M)
YORS(12) = YMU( 76,L,M)
YORS(13) = YMU( 77,L,M)
YORS(14) = YMU( 46,L,M)
YORS(15) = YMU( 31,L,M)
YORS(16) = YMU(128,L,M)
CALL YATES
DO 379 I=1,IT
B(I) = B(I)/FIT
B(LP(I,L,M)= B(I)

```

```

379 CONTINUE
  KK = KG
  DO 571 I=1,117
571 YORS(I) = 0.0
    YORS(64) = B(12)
    YORS(95) = B(13)
    YORS(96) = B(14)
    YORS(111) = B(16)
    YORS(112) = B(15)
  DO 572 I=118,128
    IPPIMI = 129-I
572 YORS(I) = B(IPPIMI)
  CALL YATES
  GO TO 600

```

```

C
208 YORS( 1) = YMU( 1,L,M)
    YORS( 2) = YMU(226,L,M)
    YORS( 3) = YMU(211,L,M)
    YORS( 4) = YMU( 52,L,M)
    YORS( 5) = YMU(181,L,M)
    YORS( 6) = YMU( 86,L,M)
    YORS( 7) = YMU(103,L,M)
    YORS( 8) = YMU(136,L,M)
    YORS( 9) = YMU(121,L,M)
    YORS(10) = YMU(154,L,M)
    YORS(11) = YMU(171,L,M)
    YORS(12) = YMU( 76,L,M)
    YORS(13) = YMU(205,L,M)
    YORS(14) = YMU( 46,L,M)
    YORS(15) = YMU( 31,L,M)
    YORS(16) = YMU(256,L,M)
  CALL YATES
  DO 389 I=1,17
    B(I) = B(I)/FIT
    RILN(I,L,M) = B(I)

```

```

389 CONTINUE
  KK = KG
  DO 581 I=1,245
581 YORS(I) = 0.0
    YORS(128) = B(8)
    YORS(192) = B(12)
    YORS(223) = B(13)
    YORS(224) = B(14)
    YORS(239) = B(16)
    YORS(240) = B(15)
  DO 582 I=246,256
    IPPIMI = 257-I
582 YORS(I) = B(IPPIMI)
    YORS(249) = 0.0
  CALL YATES
  GO TO 600

```

```

C
600 DO 609 I=1,IP
    IPPIMI = IPP1-I
    BIAS(I,L,M) = B(IPPIMI)-YMU(I,L,M)
609 CONTINUE

```

680 CONTINUE
690 CONTINUE

C
C
C

8.- OUTPUT

```

WRITE (6.830) IT,NE,KG,MP,ALPHA(KP),ALPHA(KF)
WRITE (6.832) (TAU(M),M=1,MTAU)
WRITE (6.838)
WRITE (6.839) ((THETA(L),I,(YMU(I,L,M),M=1,MTAU),I=1,IP),L=1,LTH)
WRITE (6.832) (TAU(M),M=1,MTAU)
WRITE (6.840)
WRITE (6.839) ((THETA(L),I,(BILM(I,L,M),M=1,MTAU),I=1,IT),L=1,LTH)
WRITE (6.832) (TAU(M),M=1,MTAU)
WRITE (6.841)
WRITE (6.839) ((THETA(L),I,(BIAS(I,L,M),M=1,MTAU),I=1,IP),L=1,LTH)
800 FORMAT (13A6,A2)
801 FORMAT (1H1.10X,13A6,A2//)
824 FORMAT (3I5)
825 FORMAT (15.8F5.0)
830 FORMAT (1H0.3X,4HIT =15.5X,4HNE =15.5X,4HKG =15.6X,4HMP =15.5X,
      18HALPHAP =F6.3,5X,8HALPHAF =F6.3//)
832 FORMAT (1H1.4HTAU=6F13.2//)
838 FORMAT (1H0.5HTHETA.1X.1HI.3X,3HYMU//)
839 FORMAT (1H0.F4.0,14.6E14.4)
840 FORMAT (1H0.5HTHETA.1X.1HI.3X,4HBILM//)
841 FORMAT (1H0.5HTHETA.1X.1HI.3X,4HBIAS//)
842 FORMAT (1H0.4HPSI=8E14.4)
843 FORMAT (1H0.4HPI=8E14.4)
844 FORMAT (1H0.4HDEL=8E14.4)
      ENC

```

```

SUBROUTINE YATES
COMMON KK,Y(256),B(256)
II = 2**KK
IICB2 = II/2
KKM1 = KK-1
DO 908 K=1,KKM1
DO 906 I=1,II,2
IP102 = (I+1)/2
B(IP102) = Y(I+1)+Y(I)
LL = IP102+IICB2
906 B(LL) = Y(I+1)-Y(I)
DO 907 I=1,II
907 Y(I) = B(I)
908 CONTINUE
DO 909 I=1,II,2
IP102 = (I+1)/2
B(IP102) = Y(I+1)+Y(I)
LL = IP102+IICB2
B(LL) = Y(I+1)-Y(I)
909 CONTINUE
RETURN
END

```

REFERENCES

- Allen, D. M. 1971. Mean Square Error of Prediction as a Criterion for Selecting Variables. Technometrics 13: 469-481.
- Amling, G. E., and Holms, A. G. 1973. POOLMS - A Computer Program for Fitting and Model Selection for Two-Level Factorial Replication-Free Experiments. National Aeronautics and Space Administration Tech. Memo X-2706.
- Beale, E. M. L. 1970. Note on Procedures for Variable Selection in Multiple Regression. Technometrics 12: 909-14.
- Box, G. E. P., and Hunter, J. S. 1957. Multi-Factor Experimental Designs for Exploring Response Surfaces. Ann. Math. Stat. 28: 195-241.
- Box, G. E. P., and Muller, M. E. 1958. A Note on the Generation of Random Normal Deviates. Annals of Mathematical Statistics, 29: 610-11.
- Cochran, W. G. 1941. The Distribution of the Largest of a Set of Estimated Variances as a Fraction of Their Total. Annals of Eugenics 11: 47-52.
- Collins, H. E., Quigg, R. J., and Dreshfield, R. L. 1968. Development of a Nickel-Base Superalloy Using Statistically Designed Experiments. ASM Trans. 61: 711-21.
- Daniel, C. 1959.: Use of Half-Normal Plots in Interpreting Factorial Two-Level Experiments. Technometrics 1: 311-42.
- Daniel, C., and Wood, F. S. 1971. Fitting Equations to Data. New York: Wiley Interscience.
- Draper, N. R., and Smith, H. 1966. Applied Regression Analysis. New York: John Wiley and Sons, Inc.
- Duckworth, W. E. 1965. Statistical Method in Metallurgical Development. The Statistician. 15: 7-30.
- Eckert, R. E., and Serafini, T. T. 1968. Effect of Film Processing on Cryogenic Properties of Poly(Ethylene Terephthalate). National Aeronautics and Space Administration Tech. Note D-4762.
- Efroymson, M. A. 1960 Multiple Regression Analysis. Mathematical Methods for Digital Computers, eds. A. Ralston and H. S. Wilf, pp. 191-203, New York: Wiley.
- Eiselstein, H. L. 1971. Properties of a Fabricable, High Strength Superalloy. Metals Engineering Quarterly 11: 20-5.

- Filippi, A. M. 1974. TD-NiCr (Ni-20Cr-2ThO₂) Forging Studies. Westinghouse Astronuclear Lab. Report WANL-M-FR-74-001.
- Gorman, J. W., and Toman, R. J. 1966. Selection of Variables for Fitting Equations to Data. Technometrics 8: 27-51.
- Hocking, R. R. 1972. Criteria for Selection of a Subset Regression: Which One Should be Used? Technometrics 14: 967-70.
- Hoerl, A. E., and Kennard, R. W. 1970. Ridge Regression: Biased Estimation for Nonorthogonal Problems. Technometrics 12: 55-67.
- Holms, Arthur G. 1967. Designs of Experiments as Telescoping Sequences of Blocks for Optimum Seeking (as intended for alloy development). National Aeronautics and Space Administration Tech. Note D-4100.
- Holms, A. G., and Berrettoni, J. N. 1967. Multiple Decision Procedures for ANOVA of Two-Level Factorial Fixed-Effects Replication-Free Experiments. National Aeronautics and Space Administration Tech. Note D-4272.
- Holms, A. G., and Berrettoni, J. N. 1969. Chain-Pooling ANOVA for Two-Level Factorial Replication-Free Experiments. Technometrics 11: 725-46.
- Jeffers, J. N. R. 1967. Two Case Studies in the Application of Principal Component Analysis. Applied Statistics 16: 225-36.
- Kendall, M. G., and Stuart, A. 1966. The Advanced Theory of Statistics. London: Charles Griffen & Co. Ltd.
- Kendall, M. G. 1957. A Course in Multivariate Analysis. New York: Hafner Publishing Co.
- Kennedy, W. J., and Bancroft, T. A. 1971. Model Building for Prediction in Regression Based upon Repeated Significance Tests. Ann. Math. Statistics 42: 1273-84.
- Luce, R. D., and Raiffa, H. 1957. Games and Decisions. New York: John Wiley and Sons, p. 67.
- Mallows, C. L. 1973. Some Comments on Cp. Technometrics 15: 661-675.
- Marquart, D. W. 1970. Generalized Inverses, Ridge Regression, Biased Linear Estimation, and Nonlinear Estimation. Technometrics 12: 591-612.
- Massy, W. F. 1965. Principal Components Regression in Exploratory Statistical Research. J. Am. Statist. Assoc. 60: 234-56.
- Newhouse, J. P., and Oman, S. D. 1971. An Evaluation of Ridge Estimators. Rand Corp. Report R-716-PR.

Sandrock, G. D., Holms, A. G. 1969. Statistical Design and Analysis of Optimum Seeking Experiments to Develop a Gamma-Prime Strengthened Cobalt-Nickel Base Alloy. National Aeronautics and Space Administration Tech. Note D-5587.

Sclove, S. L. 1968. Improved Estimators for Coefficients in Linear Regression. J. Am. Statist. Assoc 63: 596-606.

Sidik, S. M., and Holms, A. G. 1971. Optimal Design Procedure for Two-Level Fractional Factorial Experiments Given Prior Information about Parameters. National Aeronautics and Space Administration Tech. Note D-6527.

Sidik, S. M. 1972. An Improved Multiple Linear Regression and Data Analysis Computer Program Package. National Aeronautics and Space Administration Tech. Note D-6770.

Sidik, S. M. 1972. NAMER - A FORTRAN IV Program for use in Optimizing Designs of Two-Level Factorial Experiments Given Partial Prior Information. National Aeronautics and Space Administration Tech. Note D-6545.

Sidik, S. M. 1973. An Algorithm for Generating all Possible 2^{p-a} Fractional Factorial Designs and its use in Scientific Experimentation. National Aeronautics and Space Administration Tech. Note D-7327.

Taussky, O., and Todd, J. 1956. Generation of Pseudo-Random Numbers. In Symposium on Monte Carlo Methods, ed. H. A. Meyer, pp. 15-28, New York: Wiley.

Walls, R. C., and Weeks, D. L. 1969. A Note on the Variance of a Predicted Response in Regression. Amer. Statistician 23: 24-26.

TABLE I. - SAMPLE FRACTILES FOR ORDER STATISTICS
FROM SAMPLES OF SIZE g

$$q_j = \frac{j}{1+g} \quad j = 1, \dots, g$$

j	g				
	4	5	6	7	8
1	0.200	0.167	0.143	0.125	0.111
2	.400	.333	.286	.250	.222
3	.600	.500	.429	.375	.333
4	.800	.667	.571	.500	.444
5		.833	.714	.625	.556
6			.857	.750	.667
7				.875	.778
8					.889

TABLE II. - VALUES OF ψ_k

k	g				
	4	5	6	7	8
1	1.00	1.00	1.00	1.00	1.00
2	.65	.67	.72	.75	.76
3	.41	.45	.54	.58	.63
4	.20	.28	.39	.44	.47
5		.14	.26	.32	.37
6			.12	.20	.27
7				.10	.17
8					.07

TABLE III. - POINTS OF INFLECTION (EQ. (24))

$$[\delta_k = 1]$$

k	g				
	4	5	6	7	8
1	0.29	0.29	0.29	0.29	0.29
2	.12	-.14	-.17	.18	.19
3	-.10	-.05	-.04	.07	.11
4	-.58	-.34	-.13	-.07	-.03
5		-.89	-.39	-.25	-.16
6			-1.04	-.58	-.36
7				-1.24	-.71
8					-1.67

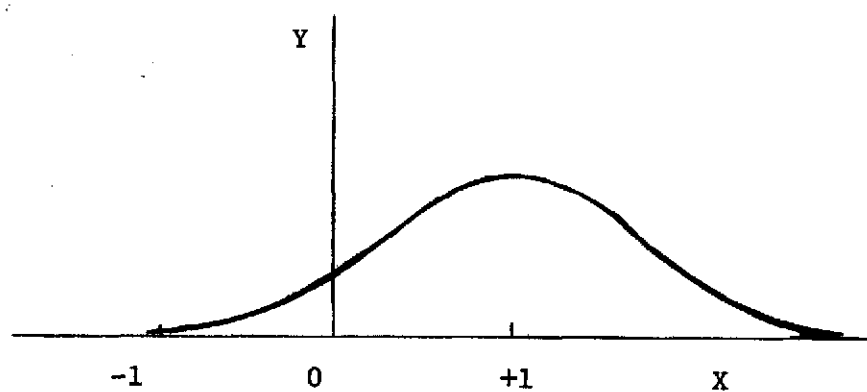


TABLE IV. - GENERATORS OF EXPERIMENT PLANS

	Full replicate estimable contrasts	Defining contrast generators			
g h	4 0	5 1	6 2	7 3	8 4
	X_0 X_1 X_2 X_1X_2 X_3 X_1X_3 X_2X_3 $X_1X_2X_3$ X_4 X_1X_4 X_2X_4 $X_1X_2X_4$ X_3X_4 $X_1X_3X_4$ $X_2X_3X_4$ $X_1X_2X_3X_4$				$X_1X_2X_3X_8$ $X_1X_2X_4X_7$ $X_1X_3X_4X_6$ $X_2X_3X_4X_5$

TABLE V. - DEFINING CONTRASTS

g			
5	6	7	8
x_0 $x_2 x_3 x_4 x_5$	x_0 $x_2 x_3 x_4 x_5$ $x_1 x_3 x_4 x_6$ $x_1 x_2 x_5 x_6$	x_0 $x_2 x_3 x_4 x_5$ $x_1 x_3 x_4 x_6$ $x_1 x_2 x_5 x_6$ $x_1 x_2 x_4 x_7$ $x_1 x_3 x_5 x_7$ $x_2 x_3 x_6 x_7$ $x_4 x_5 x_6 x_7$	x_0 $x_2 x_3 x_4 x_5$ $x_1 x_3 x_4 x_6$ $x_1 x_2 x_5 x_6$ $x_1 x_2 x_4 x_7$ $x_1 x_3 x_5 x_7$ $x_2 x_3 x_6 x_7$ $x_4 x_5 x_6 x_7$ $x_1 x_2 x_3 x_8$ $x_1 x_4 x_5 x_8$ $x_2 x_4 x_6 x_8$ $x_3 x_5 x_6 x_8$ $x_3 x_4 x_7 x_8$ $x_2 x_5 x_7 x_8$ $x_1 x_6 x_7 x_8$ $x_1 x_2 x_3 x_4 x_5 x_6 x_7 x_8$

TABLE VI. - LEVELS OF INDEPENDENT VARIABLES

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
-1	-1	-1	-1	-1	-1	-1	-1
1	-1	-1	-1	-1	1	1	1
-1	1	-1	-1	1	-1	1	1
1	1	-1	-1	1	1	-1	-1
-1	-1	1	-1	1	1	-1	1
1	-1	1	-1	1	-1	1	-1
-1	1	1	-1	-1	1	1	-1
1	1	1	-1	-1	-1	-1	1
-1	-1	-1	1	1	1	1	-1
1	-1	-1	1	1	-1	-1	1
-1	1	-1	1	-1	1	-1	1
1	1	-1	1	-1	-1	1	-1
-1	-1	1	1	-1	-1	1	1
-1	-1	1	1	-1	1	-1	-1
-1	1	1	1	1	-1	-1	-1
1	1	1	1	1	1	1	1

TABLE VII. - YATES' ORDER OF TREATMENTS

1 (1)	33 f	65 g	97 fg	129 h	161 fh	193 gh	225 fgh
2 a	34 af	66 ag	98 afg	130 ah	162 afh	194 agh	226 afgh
3 b	35 bf	67 bg	99 bfg	131 bh	163 bfh	195 bgh	227 bfigh
4 ab	36 abf	68 abg	100 abfg	132 abh	164 abfh	196 abgh	228 abfgh
5 c	37 cf	69 cg	101 cfg	133 ch	165 cfh	197 cgh	229 cfigh
6 ac	38 acf	70 acg	102 acfg	134 ach	166 acfh	198 acgh	230 acfgh
7 bc	39 bcf	71 bcg	103 bcfg	135 bch	167 bcfh	199 bcgh	231 bcfgh
8 abc	40 abcf	72 abcg	104 abcfg	136 abch	168 abcfh	200 abcdh	232 abcfgh
9 d	41 df	73 dg	105 dfg	137 dh	169 dfh	201 dgh	233 dfigh
10 ad	42 adf	74 adg	106 adfg	138 adh	170 adfh	202 adgh	234 adfgh
11 bd	43 bdf	75 bdg	107 bdfg	139 bdh	171 bdfh	203 bdgh	235 bdfgh
12 abd	44 abdf	76 abdg	108 abdfg	140 abdh	172 abdfh	204 abdgh	236 abdfgh
13 cd	45 cdf	77 cdg	109 cdfg	141 cdh	173 cdfh	205 cdgh	237 cdfgh
14 acd	46 acdf	78 acdg	110 acdfg	142 acdh	174 acdfh	206 acdgh	238 acdfgh
15 bcd	47 bcdf	79 bcdg	111 bcdfg	143 bcdh	175 bcdfh	207 bcdgh	239 bcdfgh
16 abcd	48 abcdf	80 abcdg	112 abcdfg	144 abcdh	176 abcdfh	208 abcdgh	240 abcdfigh
17 e	49 ef	81 eg	113 efg	145 eh	177 efh	209 egh	241 efigh
18 ae	50 aef	82 aeg	114 aefg	146 aeh	178 aefh	210 aegh	242 aefgh
19 be	51 bef	83 beg	115 befg	147 beh	179 befh	211 begh	243 befigh
20 abe	52 abef	84 abeg	116 abefg	148 abeh	180 abefh	212 abegh	244 abefgh
21 ce	53 cef	85 ceg	117 cefg	149 ceh	181 cefh	213 cegh	245 cefgh
22 ace	54 acef	86 aceg	118 acefg	150 aceh	182 acefh	214 acegh	246 acefgh
23 bce	55 bcef	87 bceg	119 bcefg	151 bceh	183 bcefh	215 bcegh	247 bcefgh
24 abce	56 abcef	88 abceg	120 abcefg	152 abceh	184 abcefh	216 abcegh	248 abcefigh
25 de	57 def	89 deg	121 defg	153 deh	185 defh	217 deg	249 defgh
26 ade	58 adef	90 adeg	122 adefg	154 adeh	186 adefh	218 adegh	250 adefgh
27 bde	59 bdef	91 bdeg	123 bdefg	155 bdeh	187 bdefh	219 bdegh	251 bdefgh
28 abde	60 abdef	92 abdeg	124 abdefg	156 abdeh	188 abdefh	220 abdegh	252 abdefgh
29 cde	61 cdef	93 cdeg	125 cdefg	157 cdeh	189 cdefh	221 cdegh	253 cdefgh
30 acde	62 acdef	94 acdeg	126 acdefg	158 acdeh	190 acdefh	222 acdegh	254 acdefgh
31 bcde	63 bcdef	95 bcdeg	127 bcdefg	159 bcdeh	191 bcdefh	223 bcdegh	255 bcdefgh
32 abcde	64 abcdef	96 abcdeg	128 abcdefg	160 abcdeh	192 abcdefh	224 abcdegh	256 abcdefgh

TABLE VIII. - TREATMENTS OF EXPERIMENT PLANS AND ALIASED COMBINATIONS OF LOWER ORDER PARAMETERS.

g h	4 0	5 1	6 2	7 3	8 4
1 (1)	β_0	1 (1)	β_0	1 (1)	β_0
2 a	β_1	2 a	β_1	98 afg	β_1
3 b	β_2	19 be	β_2	83 beg	β_2
4 ab	β_{12}	20 abe	β_{12}	52 abef	$\beta_{12} + \beta_{47} + \beta_{56}$
5 c	β_3	21 ce	β_3	53 cef	β_3
6 ac	β_{13}	22 ace	β_{13}	86 aceg	$\beta_{13} + \beta_{46} + \beta_{57}$
7 bc	β_{23}	7 bc	$\beta_{23} + \beta_{45}$	103 bcfg	$\beta_{23} + \beta_{45} + \beta_{67}$
8 abc	β_{123}	8 abc	$\beta_{123} + \beta_{145}$	8 abc	$\beta_{123} + \beta_{145} + \beta_{246} + \beta_{356} + \beta_{257} + \beta_{167}$
9 d	β_4	25 de	β_4	121 defg	β_4
10 ad	β_{14}	26 ade	β_{14}	26 ade	$\beta_{14} + \beta_{36} + \beta_{27}$
11 bd	β_{24}	11 bd	$\beta_{24} + \beta_{35}$	43 bdf	$\beta_{24} + \beta_{35} + \beta_{17}$
12 abd	β_{124}	12 abd	$\beta_{124} + \beta_{135}$	76 abd	$\beta_{124} + \beta_7$
13 cd	β_{34}	13 cd	$\beta_{34} + \beta_{25}$	77 cdg	$\beta_{34} + \beta_{25} + \beta_{16}$
14 acd	β_{134}	14 acd	$\beta_{134} + \beta_{125}$	46 acdf	$\beta_{134} + \beta_6$
15 bcd	β_{234}	31 bcde	$\beta_{234} + \beta_5$	31 bcde	$\beta_{234} + \beta_5$
16 abcd	β_{1234}	32 abcde	$\beta_{1234} + \beta_{15}$	128 abcdefg	$\beta_{1234} + \beta_{15} + \beta_{26} + \beta_{37}$
					256 abcdefgh
					$\beta_{1234} + \beta_{15} + \beta_{26} + \beta_{37} + \beta_{48}$

Numbers in columns are the Yates order number of the treatment in a full factorial experiment, as in table VII

TABLE IX. - STARTING MODELS FOR SUBSET REGRESSION PROCEDURES

[The number i is the order number of the Yates estimate for a full factorial experiment.]

$g = 4, h = 0$															
i															
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
$y = \beta_0$	$+\beta_1 x_1$	$+\beta_2 x_2$	$+\beta_{12} x_1 x_2$	$+\beta_3 x_3$	$+\beta_{13} x_1 x_3$	$+\beta_{23} x_2 x_3$	$+\beta_{123} x_1 x_2 x_3$	$+\beta_4 x_4$	$+\beta_{14} x_1 x_4$	$+\beta_{24} x_2 x_4$	$+\beta_{124} x_1 x_2 x_4$	$+\beta_{34} x_3 x_4$	$+\beta_{134} x_1 x_3 x_4$	$+\beta_{234} x_2 x_3 x_4$	$+\beta_{1234} x_1 x_2 x_3 x_4$
$g = 5, h = 1$															
i															
1	2	3	4	5	6	7	8	9	10	11	12	19	20	17	18
$y = \beta_0$	$+\beta_1 x_1$	$+\beta_2 x_2$	$+\beta_{12} x_1 x_2$	$+\beta_3 x_3$	$+\beta_{13} x_1 x_3$	$+\beta_{23} x_2 x_3$	$+\beta_{123} x_1 x_2 x_3$	$+\beta_4 x_4$	$+\beta_{14} x_1 x_4$	$+\beta_{24} x_2 x_4$	$+\beta_{124} x_1 x_2 x_4$	$+\beta_{25} x_2 x_5$	$+\beta_{125} x_1 x_2 x_5$	$+\beta_5 x_5$	$+\beta_{15} x_1 x_5$
$g = 6, h = 2$															
i															
1	2	3	4	5	6	7	8	9	10	11	12	34	33	17	18
$y = \beta_0$	$+\beta_1 x_1$	$+\beta_2 x_2$	$+\beta_{12} x_1 x_2$	$+\beta_3 x_3$	$+\beta_{13} x_1 x_3$	$+\beta_{23} x_2 x_3$	$+\beta_{123} x_1 x_2 x_3$	$+\beta_4 x_4$	$+\beta_{14} x_1 x_4$	$+\beta_{24} x_2 x_4$	$+\beta_{124} x_1 x_2 x_4$	$+\beta_{16} x_1 x_6$	$+\beta_6 x_6$	$+\beta_5 x_5$	$+\beta_{15} x_1 x_5$
$g = 7, h = 3$															
i															
1	2	3	4	5	6	7	8	9	10	11	65	34	33	17	18
$y = \beta_0$	$+\beta_1 x_1$	$+\beta_2 x_2$	$+\beta_{12} x_1 x_2$	$+\beta_3 x_3$	$+\beta_{13} x_1 x_3$	$+\beta_{23} x_2 x_3$	$+\beta_{123} x_1 x_2 x_3$	$+\beta_4 x_4$	$+\beta_{14} x_1 x_4$	$+\beta_{24} x_2 x_4$	$+\beta_7 x_7$	$+\beta_{16} x_1 x_6$	$+\beta_6 x_6$	$+\beta_5 x_5$	$+\beta_{15} x_1 x_5$
$g = 8, h = 4$															
i															
1	2	3	4	5	6	7	129	9	10	11	65	34	33	17	18
$y = \beta_0$	$+\beta_1 x_1$	$+\beta_2 x_2$	$+\beta_{12} x_1 x_2$	$+\beta_3 x_3$	$+\beta_{13} x_1 x_3$	$+\beta_{23} x_2 x_3$	$+\beta_8 x_8$	$+\beta_4 x_4$	$+\beta_{14} x_1 x_4$	$+\beta_{24} x_2 x_4$	$+\beta_7 x_7$	$+\beta_{16} x_1 x_6$	$+\beta_6 x_6$	$+\beta_5 x_5$	$+\beta_{15} x_1 x_5$

TABLE X. - COMPUTER PROGRAMS

Name	σ^2	m_p	α_p	α_f	n_e	Output
MODEL	0	0	1.0	1.0		YMU(I,L,M) I = 1,IP BILM(I,L,M) I = 1,IT BIAS(I,L,M) I = 1,IP
POOLES	0	Optional				ERSQMX (L,M) AVERSQ (L,M)
POOLES	1	Optional			Optional	ERSQMX (L,M) AVERSQ (L,M) VESQMX (L,M) AVVESQ (L,M)

TABLE XI. - STRATEGIES OF α_p AND α_f
(66 COMBINATIONS)

α_p	α_f	α_p	α_f	α_p	α_f
1.0	1.0	0.25	0.25	0.005	0.005
	.75		.10		.002
	.50		.05		.001
	.25		.025		.002
	.10		.01		.001
	.05		.005		.001
	.025		.002		
	.01		.001		
	.005	.10	.10	.002	
	.002		.05		
	.001		.025		
.75	.75	.05	.01	.001	
	.50		.005		
	.25		.002		
	.10		.001		
	.05		.05		
	.025		.025		
	.01		.01		
	.005	.025	.005	.002	
	.002		.002		
	.001		.001		
.50	.50	.01	.025	.001	
	.25		.01		
	.10		.005		
	.05		.002		
	.025		.001		
	.01		.01		
	.005	.025	.005	.002	
	.002		.002		
	.001		.001		

TABLE XII. - SUMMARY OF POPULATION CHARACTERISTICS AS
DETERMINED WITH PROGRAM MODEL

[Values of i are Yates order number of treatment
at which maximum absolute values of mean and
bias occurred.]

g	h	τ	θ							
			1				32			
			$ \mu_{1m} _{\max}$	i	$ \text{bias} _{\max}$	i	$ \mu_{1m} _{\max}$	i	$ \text{bias} _{\max}$	i
4	0	0	3.260	16	0.0	-	104.3	16	0.0	-
		1	3.262	16	0.0	-	104.4	16	0.0	-
		2	3.263	16	0.0	-	104.4	16	0.0	-
		4	4.040	3	0.0	-	129.3	3	0.0	-
		8	8.040	3	0.0	-	257.3	3	0.0	-
		16	16.04	3	0.0	-	513.3	3	0.0	-
5	1	0	3.540	32	0.0	-	113.3	32	0.0	-
		1	3.541	32	.4840	3	113.3	32	15.49	3
		2	3.541	32	.9680	3	113.3	32	30.98	3
		4	3.800	3	1.936	3	121.6	3	61.95	3
		8	7.800	3	3.872	3	249.6	3	123.9	3
		16	15.80	3	7.744	3	505.6	3	247.8	3
6	2	0	4.030	64	0.0	-	129.0	64	0.0	-
		1	4.030	64	.6733	3	129.0	64	21.55	3
		2	4.030	64	1.347	3	129.0	64	43.09	3
		4	4.030	64	2.693	3	129.0	64	86.19	3
		8	7.410	3	5.387	3	237.1	3	172.4	3
		16	15.41	3	10.77	3	493.1	3	344.7	3
7	3	0	4.390	128	0.0	-	140.5	128	0.0	-
		1	4.390	128	.8328	3	140.5	128	26.65	3
		2	4.390	128	1.666	3	140.5	128	53.30	3
		4	4.390	128	3.331	3	140.5	128	106.6	3
		8	7.110	3	6.662	3	227.5	3	213.2	3
		16	15.11	3	13.32	3	483.5	3	426.4	3
8	4	0	4.740	256	0.0	-	151.7	256	0.0	-
		1	4.740	256	.9191	3	151.7	256	29.41	3
		2	4.740	256	1.838	3	151.7	256	58.82	3
		4	4.740	256	3.676	3	151.7	256	117.6	3
		8	6.780	3	7.352	3	217.0	3	235.3	3
		16	14.78	3	14.70	3	473.0	3	470.6	3

$\delta_k = -1, 1, -1, -1, \dots$

$k = 1, \dots, g$

$$\mu_{1m} = \theta_1 \left[1 + \sum_{k=1}^g \psi_k x_{1k} + \tau_m e^{-\sum_{k=1}^g \left(\frac{x_{1k} - \delta_k}{\psi_k^{-1/2}} \right)^2} \right] \quad (26)$$

TABLE XIII. - OPTIMAL STRATEGIES FOR GIVEN θ AND τ (a) $\tau = 0.0$; $\delta_k = -1, 1, -1, -1, \dots$; $n_e = 1000$

θ	g	4	5	6	7	8
	h	0	1	2	3	4
0.125	α_p	1.00	1.00	1.00	1.00	1.00
	α_f	.001	.001	.001	.001	.001
	\bar{e}_{\max}^2	.1444	.1656	.2086	.2450	.2843
	$V(\bar{e}^2)_{\max}$.0302	.0357	.0467	.0560	.0660
.25	α_p	1.00	1.00	.002	.25	.50
	α_f	.001	.001	.002	.001	.005
	\bar{e}_{\max}^2	.3855	.4700	.6411	.7564	.8794
	$V(\bar{e}^2)_{\max}$.0918	.1133	.1577	.3418	.9264
.5	α_p	.75	.75	.75	1.00	^a 1.00
	α_f	.05	.05	.50	.25	1.00
	\bar{e}_{\max}^2	.9423	1.019	1.092	1.091	1.117
	$V(\bar{e}^2)_{\max}$	1.616	1.645	2.545	2.551	2.584
1.0	α_p	.75	.75	1.00	1.00	1.00
	α_f	.10	.10	.25	.25	.50
	\bar{e}_{\max}^2	.9798	1.033	1.100	1.114	1.113
	$V(\bar{e}^2)_{\max}$	1.973	2.054	2.436	2.461	2.612
2.0	α_p	.75	.75	1.00	1.00	.75
	α_f	.10	.10	.25	.25	.75
	\bar{e}_{\max}^2	.9559	.9912	1.057	1.108	1.113
	$V(\bar{e}^2)_{\max}$	1.859	2.128	2.450	2.482	2.536
4.0	α_p	.10	.75	.75	.50	1.00
	α_f	.10	.10	.10	.10	.25
	\bar{e}_{\max}^2	.7964	.9573	.9769	1.056	1.069
	$V(\bar{e}^2)_{\max}$	1.407	2.006	2.094	2.446	2.542
8.0	α_p	.025	.10	.10	.50	.25
	α_f	.025	.10	.10	.05	.25
	\bar{e}_{\max}^2	.4786	.7731	.8800	.9458	1.016
	$V(\bar{e}^2)_{\max}$.7771	1.458	1.655	2.060	2.444
16.0	α_p	.005	.01	.025	.05	.50
	α_f	.005	.01	.025	.05	.05
	\bar{e}_{\max}^2	.3447	.4337	.5642	.7135	.9346
	$V(\bar{e}^2)_{\max}$.2322	.4721	.7246	1.356	2.002
32.0	α_p	.005	.01	.001	.005	.025
	α_f	.005	.002	.001	.002	.025
	\bar{e}_{\max}^2	.3447	.4177	.5038	.5745	.6793
	$V(\bar{e}^2)_{\max}$.2322	.3717	.4920	.6628	.9404

^a $m_p = 0$. In all other cases, $m_p = 1$.

TABLE XIII. - Continued. OPTIMAL STRATEGIES FOR GIVEN θ AND τ

(b) $\tau = 1.0$, $\delta_k = -1, 1, -1, -1, \dots$; $n_e = 1000$

θ	g_h	4 0	5 1	6 2	7 3	8 4
0.125	α_p	1.00	1.00	1.00	1.00	1.00
	α_f	.001	.001	.001	.001	.001
	\bar{e}_{\max}^2	.1441	.1624	.2049	.2427	.2828
	$V(\bar{e}^2)_{\max}$.0310	.0358	.0463	.0552	.0649
.25	α_p	.002	.002	.002	.25	.25
	α_f	.002	.002	.001	.001	.005
	\bar{e}_{\max}^2	.3883	.4623	.6257	.7494	.8850
	$V(\bar{e}^2)_{\max}$.1184	.1139	.1566	.4531	.8731
.5	α_p	.75	.50	.75	1.00	1.00
	α_f	.05	.05	.25	.25	.25
	\bar{e}_{\max}^2	.9296	.9879	1.078	1.088	1.149
	$V(\bar{e}^2)_{\max}$	1.617	1.671	2.511	2.544	2.499
1.0	α_p	.75	.75	.75	.75	.75
	α_f	.10	.10	.10	.10	.10
	\bar{e}_{\max}^2	1.003	1.076	1.218	1.314	1.547
	$V(\bar{e}^2)_{\max}$	2.224	2.196	2.892	3.984	7.046
2.0	α_p	1.00	.25	.25	.25	.50
	α_f	.25	.25	.10	.25	.05
	\bar{e}_{\max}^2	1.068	1.501	2.188	2.810	3.106
	$V(\bar{e}^2)_{\max}$	2.268	3.832	17.83	22.51	19.91
4.0	α_p	.75	.10	.10	.25	.25
	α_f	.75	.05	.01	.01	.01
	\bar{e}_{\max}^2	1.112	4.118	6.206	8.287	9.033
	$V(\bar{e}^2)_{\max}$	2.267	23.59	189.6	160.0	379.3
8.0	α_p	^a 1.00	.10	.025	.05	.10
	α_f	1.00	.002	.002	.002	.005
	\bar{e}_{\max}^2	1.116	15.08	24.99	29.08	39.62
	$V(\bar{e}^2)_{\max}$	2.282	198.1	491.4 $\times 10$	232.2 $\times 10$	342.4
16.0	α_p	^a 1.00	.025	.005	1.00	.025
	α_f	1.00	.002	.001	.025	.001
	\bar{e}_{\max}^2	1.116	56.11	108.0	159.8	177.9
	$V(\bar{e}^2)_{\max}$	2.282	257.9 $\times 10$	160.4 $\times 10^3$	934.1 $\times 10$	305.4 $\times 10^3$
32.0	α_p	^a 1.00	1.00	^a 1.00	1.00	.005
	α_f	1.00	.75	1.00	.025	.001
	\bar{e}_{\max}^2	1.116	236.4	462.8	681.0	842.0
	$V(\bar{e}^2)_{\max}$	2.282	974.1	196.4 $\times 10$	439.2 $\times 10^2$	258.8 $\times 10^4$

^a $m_p = 0$. In all other cases, $m_p = 1$.

TABLE XIII. - Continued. OPTIMAL STRATEGIES FOR GIVEN θ AND τ (c) $\tau = 2.0$, $\delta_k = -1, 1, -1, -1, \dots$; $n_e = 1000$

θ		4 0	5 1	6 2	7 3	8 4
0.125	α_p	1.00	1.00	1.00	1.00	1.00
	α_f	.001	.001	.001	.001	.001
	\bar{e}_{\max}^2	.1473	.1629	.2012	.2403	.2814
	$V(\bar{e}^2)_{\max}$.0319	.0360	.0459	.0545	.0639
.25	α_p	.001	.005	.005	.25	.25
	α_f	.001	.005	.005	.001	.002
	\bar{e}_{\max}^2	.4037	.4643	.6146	.7412	.8586
	$V(\bar{e}^2)_{\max}$.0984	.1571	.1646	.4785	.6676
.5	α_p	.75	1.00	1.00	.50	.50
	α_f	.05	.10	.10	.05	.10
	\bar{e}_{\max}^2	.9009	.9760	1.150	1.257	1.412
	$V(\bar{e}^2)_{\max}$	1.978	1.998	3.018	2.859	3.550
1.0	α_p	.75	.75	.75	.25	.75
	α_f	.10	.05	.05	.25	.05
	\bar{e}_{\max}^2	1.015	1.556	2.035	2.542	2.726
	$V(\bar{e}^2)_{\max}$	2.026	3.788	6.866	12.25	17.94
2.0	α_p	1.00	.10	.25	1.00	.25
	α_f	1.00	.10	.025	.05	.025
	\bar{e}_{\max}^2	1.114	3.618	6.315	7.691	8.787
	$V(\bar{e}^2)_{\max}$	2.294	26.72	51.36	143.6	210.8
4.0	α_p	^a 1.00	.05	.10	.10	.10
	α_f	1.00	.025	.005	.005	.01
	\bar{e}_{\max}^2	1.116	12.27	22.64	26.16	32.25
	$V(\bar{e}^2)_{\max}$	2.282	52.58	808.4	135.5 $\times 10$	371.1 $\times 10$
8.0	α_p	^a 1.00	.10	.025	.05	.05
	α_f	1.00	.005	.002	.001	.002
	\bar{e}_{\max}^2	1.116	47.84	100.4	105.9	144.8
	$V(\bar{e}^2)_{\max}$	2.282	801.8	369.5 $\times 10^2$	292.9 $\times 10^2$	503.2 $\times 10^2$
16.0	α_p	^a 1.00	.10	.025	.025	.025
	α_f	1.00	.025	.001	.001	.001
	\bar{e}_{\max}^2	1.116	192.9	391.2	488.5	714.0
	$V(\bar{e}^2)_{\max}$	2.282	984.2 $\times 10$	117.8 $\times 10^6$	473.3 $\times 10^3$	587.3 $\times 10^3$
32.0	α_p	^a 1.00	.05	.005	1.00	.005
	α_f	1.00	.05	.001	.025	.002
	\bar{e}_{\max}^2	1.116	819.5	159.9 $\times 10$	228.0 $\times 10$	272.5 $\times 10$
	$V(\bar{e}^2)_{\max}$	2.282	112.4 $\times 10^3$	587.7 $\times 10^4$	101.8 $\times 10^4$	308.0 $\times 10^5$

^a $m_p = 0$. In all other cases $m_p = 1$.

TABLE XIII. - Continued. OPTIMAL STRATEGIES FOR GIVEN θ AND τ (d) $\tau = 0.0$; $\delta_k = 1, 1, 1, 1, \dots$; $n_e = 1000$

		4	5	6	7	8
		0	1	2	3	4
0.125	α_p	1.00	1.00	1.00	1.00	1.00
	α_f	.001	.001	.001	.001	.001
	\bar{e}_{\max}^2	.1444	.1656	.2086	.2450	.2843
	$V(\bar{e}^2)_{\max}$.0302	.0357	.0467	.0560	.0660
.25	α_p	1.00	1.00	.002	.25	.25
	α_f	.001	.001	.002	.001	.002
	\bar{e}_{\max}^2	.3855	.4700	.6411	.7564	.8930
	$V(\bar{e}^2)_{\max}$.0918	.1133	.1577	.3418	.6352
.5	α_p	.75	.75	.75	1.00	^a 1.00
	α_f	.05	.05	.50	.25	1.00
	\bar{e}_{\max}^2	.9423	1.019	1.092	1.091	1.117
	$V(\bar{e}^2)_{\max}$	1.616	1.645	2.545	2.551	2.584
1.0	α_p	.75	.75	1.00	1.00	1.00
	α_f	.10	.10	.25	.25	.50
	\bar{e}_{\max}^2	.9798	1.033	1.100	1.114	1.113
	$V(\bar{e}^2)_{\max}$	1.973	2.054	2.436	2.461	2.612
2.0	α_p	.75	.75	1.00	1.00	.75
	α_f	.10	.10	.25	.25	.75
	\bar{e}_{\max}^2	.9559	.9912	1.057	1.108	1.113
	$V(\bar{e}^2)_{\max}$	1.859	2.128	2.450	2.482	2.536
4.0	α_p	.10	.75	.75	.50	1.00
	α_f	.10	.10	.10	.10	.25
	\bar{e}_{\max}^2	.7964	.9573	.9769	1.056	1.069
	$V(\bar{e}^2)_{\max}$	1.407	2.006	2.094	2.446	2.542
8.0	α_p	.025	.10	.10	.50	.25
	α_f	.025	.10	.10	.05	.25
	\bar{e}_{\max}^2	.4786	.7731	.8800	.9458	1.016
	$V(\bar{e}^2)_{\max}$.7771	1.458	1.655	2.060	2.444
16.0	α_p	.005	.01	.025	.05	.50
	α_f	.005	.01	.025	.05	.05
	\bar{e}_{\max}^2	.3447	.4337	.5642	.7135	.9346
	$V(\bar{e}^2)_{\max}$.2322	.4721	.7246	1.356	2.002
32.0	α_p	.005	.01	.001	.005	.025
	α_f	.005	.002	.001	.002	.025
	\bar{e}_{\max}^2	.3447	.4177	.5038	.5745	.6793
	$V(\bar{e}^2)_{\max}$.2322	.3717	.4920	.6628	.9404

^a $m_p = 0$. In all other cases $m_p = 1$.

TABLE XIII. - Continued. OPTIMAL STRATEGIES FOR GIVEN θ AND τ

(ϵ) $\tau = 1.0$; $\delta_k = 1, 1, 1, 1, \dots$; $n_e = 1000$

θ		4 0	5 1	6 2	7 3	8 4
0.125	α_p	1.00	1.00	0.001	0.001	0.001
	α_f	.001	.001	.001	.001	.001
	\bar{e}_{\max}^2	.2195	.2515	.3106	.3580	.4078
	$V(\bar{e}^2)_{\max}$.0460	.0538	.0684	.0802	.0925
.25	α_p	.75	.25	.50	.75	.75
	α_f	.001	.001	.025	.05	.05
	\bar{e}_{\max}^2	.6803	.7942	.9589	.9926	1.033
	$V(\bar{e}^2)_{\max}$.2085	.3667	1.541	1.648	1.760
.5	α_p	.75	1.00	1.00	1.00	1.00
	α_f	.10	.25	.25	.25	.25
	\bar{e}_{\max}^2	1.039	1.076	1.100	1.110	1.118
	$V(\bar{e}^2)_{\max}$	1.944	2.273	2.450	2.413	2.466
1.0	α_p	1.00	1.00	1.00	1.00	1.00
	α_f	.25	.25	.25	.25	.25
	\bar{e}_{\max}^2	1.060	1.124	1.189	1.192	1.247
	$V(\bar{e}^2)_{\max}$	2.037	2.394	2.620	2.709	3.541
2.0	α_p	1.00	1.00	1.00	1.00	1.00
	α_f	.25	.25	.25	.25	.25
	\bar{e}_{\max}^2	1.061	1.403	1.641	1.662	1.873
	$V(\bar{e}^2)_{\max}$	2.096	3.760	5.096	5.061	5.615
4.0	α_p	1.00	1.00	1.00	1.00	1.00
	α_f	.50	.25	.25	.25	.25
	\bar{e}_{\max}^2	1.090	2.401	3.180	3.278	4.327
	$V(\bar{e}^2)_{\max}$	2.332	8.209	11.79	12.82	19.62
8.0	α_p	^a 1.00	.75	.75	.75	.75
	α_f	1.00	.50	.50	.50	.25
	\bar{e}_{\max}^2	1.116	6.286	8.975	9.780	11.11
	$V(\bar{e}^2)_{\max}$	2.282	64.48	101.4	94.77	143.1
16.0	α_p	^a 1.00	1.00	1.00	1.00	1.00
	α_f	1.00	.75	.75	.75	.75
	\bar{e}_{\max}^2	1.116	17.78	28.64	35.43	45.40
	$V(\bar{e}^2)_{\max}$	2.282	150.5	369.1	478.5 $\times 10$	451.9
32.0	α_p	^a 1.00	1.00	1.00	1.00	1.00
	α_f	1.00	1.00	1.00	1.00	.75
	\bar{e}_{\max}^2	1.116	107.9	176.0	202.3	169.5
	$V(\bar{e}^2)_{\max}$	2.282	763.0	242.3 $\times 10$	350.7 $\times 10$	262.8 $\times 10$

^a $m_p = 0$. In all other cases $m_p = 1$.

TABLE XIII. - Concluded. OPTIMAL STRATEGIES FOR GIVEN θ AND τ

(f) $\tau = 2.0$; $\delta_k = 1, 1, 1, 1, 1, \dots$; $n_e = 1000$

θ		4	5	6	7	8
		0	1	2	3	4
0.125	α_p	0.002	0.001	1.00	1.00	1.00
	α_f	.001	.001	.001	.001	.001
	\bar{e}_{\max}^2	.3188	.3631	.4398	.4988	.5592
	$V(\bar{e}^2)_{\max}$.0705	.0814	.1002	.1148	.1298
.25	α_p	.50	1.00	.75	.75	1.00
	α_f	.025	.10	.10	.10	.25
	\bar{e}_{\max}^2	.9223	.9629	1.018	1.030	1.089
	$V(\bar{e}^2)_{\max}$	1.381	1.840	2.109	2.150	2.448
.5	α_p	1.00	1.00	1.00	.75	1.00
	α_f	.25	.25	.25	.25	.25
	\bar{e}_{\max}^2	1.048	1.132	1.191	1.220	1.216
	$V(\bar{e}^2)_{\max}$	2.080	2.382	2.594	2.823	2.958
1.0	α_p	1.00	1.00	1.00	1.00	1.00
	α_f	.25	.25	.25	.25	.25
	\bar{e}_{\max}^2	1.059	1.414	1.669	1.716	1.842
	$V(\bar{e}^2)_{\max}$	2.215	3.851	5.150	5.152	5.435

OF α_P AND α_F WERE OPTIMAL

[illegible]

^dSee table XV(d).

TABLE XIV. - Concluded. VALUES OF g AND θ AT WHICH STATEDVALUES OF α_p AND α_f WERE OPTIMAL(b) $\tau = 1.0$; $\delta_k = 1, 1, 1, 1, \dots$; $n_e = 1000$

α_p	α_f										
	1.00	0.75	0.50	0.25	0.10	0.05	0.025	0.01	0.005	0.002	0.001
^a 1.00	4,8.0 4,16.0 ^g 4,32.0										
1.00	5,32.0 6,32.0 7,32.0	5,16.0 6,16.0 7,16.0 8,32.0	4,4.0 8,16.0	^f 4,1.0 4,2.0 5,0.5 5,1.0 5,2.0 5,4.0 6,0.5 6,1.0 6,2.0 6,4.0 7,0.5 7,1.0 7,2.0 7,4.0 8,0.5 8,1.0 8,2.0 8,4.0							^e 4,0.125 5,0.125
.75			5,8.0 6,8.0 7,8.0	8,8.0	4,0.5	7,0.25 8,0.25					4,0.25
.50							6,0.25				
.25											5,0.25
.001											6,0.125 7,0.125 ^h 8,0.125

^eSee table XV(e).^fSee table XV(f).^gSee table XV(g).^hSee table XV(h).

TABLE XV. - EFFECT OF α_p AND α_f ON MEAN SQUARED ERROR AND ON VARIANCE OF MEAN SQUARED ERROR(a) $g = 4$; $h = 0$; $\theta = 0.125$; $\tau = 0.0$; $\delta_k = -1, 1, -1, -1$, and $\delta_k = 1, 1, 1, 1$; $n_e = 1000$

α_p		α_f										
		1.00	0.75	0.50	0.25	0.10	0.05	0.025	0.01	0.005	0.002	0.001
1.00	\bar{e}_{\max}^2	1.116										
	$V(\bar{e}^2)_{\max}$	2.282										
1.00	\bar{e}_{\max}^2	1.116	1.112	1.103	1.043	0.8372	0.6010	0.4108	0.2738	0.2190	0.1922	0.1444
	$V(\bar{e}^2)_{\max}$	2.275	2.294	2.279	2.091	1.599	1.155	.6710	.4263	.2155	.1681	.0302
.75	\bar{e}_{\max}^2		1.091	1.085	1.050	0.9391	.7818	.5751	.3835	.3059	.2525	.1902
	$V(\bar{e}^2)_{\max}$		2.246	2.251	2.173	1.881	1.530	1.105	.6344	.4867	.2804	.1860
.50	\bar{e}_{\max}^2			1.020	.9840	.9194	.8250	.6943	.5157	.4000	.3195	.2487
	$V(\bar{e}^2)_{\max}$			2.238	2.162	1.932	1.653	1.323	.9057	.6548	.4335	.3008
.25	\bar{e}_{\max}^2				.7182	.6618	.6216	.5810	.4693	.3887	.3360	.2646
	$V(\bar{e}^2)_{\max}$				1.544	1.467	1.364	1.146	.9791	.7019	.5091	.3498
.10	\bar{e}_{\max}^2					.4100	.3867	.3750	.3429	.3169	.2794	.2284
	$V(\bar{e}^2)_{\max}$.8122	.7263	.6804	.5710	.5168	.4220	.2612
.05	\bar{e}_{\max}^2						.2906	.2815	.2602	.2472	.2385	.2077
	$V(\bar{e}^2)_{\max}$.4488	.4109	.3647	.3364	.2905	.2180
.025	\bar{e}_{\max}^2							.2230	.2135	.2069	.2037	.1772
	$V(\bar{e}^2)_{\max}$.2712	.2530	.2364	.2547	.1975
.01	\bar{e}_{\max}^2								.1855	.1827	.1823	.1629
	$V(\bar{e}^2)_{\max}$.1613	.1413	.1448	.1267
.005	\bar{e}_{\max}^2									.1637	.1634	.1555
	$V(\bar{e}^2)_{\max}$.1135	.1057	.1209
.002	\bar{e}_{\max}^2										.1561	.1504
	$V(\bar{e}^2)_{\max}$.1038	.1143
.001	\bar{e}_{\max}^2											.1465
	$V(\bar{e}^2)_{\max}$.0325

^a $m_p = 0$. In all other cases $m_p = 1$.

TABLE XV. - Continued. EFFECT OF α_p AND α_f ON MEAN SQUARED
ERROR AND ON VARIANCE OF MEAN SQUARED ERROR

(b) $\delta_k = -1, 1, -1, -1, \dots$; $\delta_k = 1, 1, 1, 1, \dots$; $g = 4$; $h = 0$; $n_e = 1000$; $\theta = 1.0$; $\tau = 0.0$

α_p	α_f										
	1.00	0.75	0.50	0.25	0.10	0.05	0.025	0.01	0.005	0.002	0.001
^a 1.00 \bar{e}_{\max}^2	1.116										
$V(\bar{e}^2)_{\max}$	2.282										
1.00 \bar{e}_{\max}^2	1.115	1.116	1.109	1.042	1.090	1.719	2.829	3.987	4.557	4.792	5.184
$V(\bar{e}^2)_{\max}$	2.282	2.305	2.290	2.138	2.345	4.149	5.757	5.068	3.617	2.812	1.306
.75 \bar{e}_{\max}^2		1.101	1.088	1.044	1.9798	1.181	1.804	2.846	3.512	4.018	4.555
$V(\bar{e}^2)_{\max}$		2.244	2.263	2.214	1.973	2.574	4.404	5.831	5.659	4.860	3.597
.50 \bar{e}_{\max}^2			1.034	1.053	1.100	1.150	1.410	2.129	2.684	3.224	3.815
$V(\bar{e}^2)_{\max}$			2.191	2.167	2.169	2.447	3.446	5.075	5.795	1.254	5.321
.25 \bar{e}_{\max}^2				1.204	1.430	1.643	1.858	2.213	2.619	3.007	3.532
$V(\bar{e}^2)_{\max}$				2.470	3.247	3.868	4.375	5.138	5.704	1.212	5.648
.10 \bar{e}_{\max}^2					2.194	2.585	2.896	3.228	3.524	3.768	4.168
$V(\bar{e}^2)_{\max}$					4.994	5.577	5.664	5.822	5.718	5.459	4.691
.05 \bar{e}_{\max}^2						3.072	3.402	3.766	3.993	4.219	4.523
$V(\bar{e}^2)_{\max}$						5.879	5.618	5.311	5.002	4.551	3.704
.025 \bar{e}_{\max}^2							3.834	4.203	4.425	4.617	4.813
$V(\bar{e}^2)_{\max}$							5.203	4.458	4.003	3.504	2.818
.01 \bar{e}_{\max}^2								4.448	4.670	4.839	4.968
$V(\bar{e}^2)_{\max}$								3.863	3.289	2.748	2.248
.005 \bar{e}_{\max}^2									4.769	4.929	5.025
$V(\bar{e}^2)_{\max}$									2.943	2.382	2.003
.002 \bar{e}_{\max}^2										4.968	5.065
$V(\bar{e}^2)_{\max}$										2.220	1.831
.001 \bar{e}_{\max}^2											5.086
$V(\bar{e}^2)_{\max}$											1.747

^a $m_p = 0$. In all other cases, $m_p = 1$.

TABLE XV. - Continued.. EFFECT OF α_p AND α_f ON MEAN SQUARED ERROR AND ON VARIANCE OF MEAN SQUARED ERROR

(c) $g = 4$; $h = 0$; $\theta = 32.0$; $\tau = 0.0$; $\delta_k = -1, 1, -1, -1$; and $\delta_k = 1, 1, 1, 1$; $n_e = 1000$

α_p	α_f										
	1.00	0.75	0.50	0.25	0.10	0.05	0.025	0.01	0.005	0.002	0.001
1.00 \bar{e}_{\max}^2	1.116										
$V(\bar{e}^2)_{\max}$	2.282										
1.00 \bar{e}_{\max}^2	1.115	1.114	1.099	1.024	0.8100	0.6388	0.6226	8.046	1.222×10^2	5.293×10^2	5.231×10^3
$V(\bar{e}^2)_{\max}$	2.277	2.297	2.284	2.127	1.463	1.126	6.907	1.478×10^3	1.967×10^5	1.468×10^6	1.397×10^3
.75 \bar{e}_{\max}^2		1.096	1.080	1.031	$\boxed{.8974}$.7814	.6200	1.413	2.238×10	1.232×10^2	3.594×10^3
$V(\bar{e}^2)_{\max}$		2.250	2.247	2.233	1.893	1.510	1.040	4.169×10	1.430×10^4	2.252×10^5	5.890×10^6
.50 \bar{e}_{\max}^2			1.025	1.001	.9076	.8203	.7104	.6322	3.494	1.907×10	2.265×10^3
$V(\bar{e}^2)_{\max}$			2.200	2.200	1.921	1.606	1.201	5.160	5.497×10^2	1.367×10^4	6.723×10^6
.25 \bar{e}_{\max}^2				.8417	.7824	.7362	.6677	.5928	.5934	1.140	1.125×10^3
$V(\bar{e}^2)_{\max}$				1.664	1.560	1.386	1.152	.8564	5.120	3.297×10	4.622×10^6
.10 \bar{e}_{\max}^2					.5849	.5632	.5361	.5130	.4608	.4174	4.346×10^2
$V(\bar{e}^2)_{\max}$					1.068	.8626	.7692	.6844	.5691	.4403	2.085×10^6
.05 \bar{e}_{\max}^2						.4589	.4448	.4312	.4086	.3901	2.514×10^2
$V(\bar{e}^2)_{\max}$.5630	.5233	.4554	.4123	.3769	1.252×10^6
.025 \bar{e}_{\max}^2							.4032	.3974	.3815	.3739	1.102×10^2
$V(\bar{e}^2)_{\max}$.4001	.3854	.3591	.3335	5.630×10^5
.01 \bar{e}_{\max}^2								.3643	.3552	.3535	3.698×10
$V(\bar{e}^2)_{\max}$.2951	.2784	.2780	1.909×10^5
.005 \bar{e}_{\max}^2									$\boxed{.3447}$.3450	1.078×10
$V(\bar{e}^2)_{\max}$.2322	.2329	5.484×10^4
.002 \bar{e}_{\max}^2										.3458	1.078×10
$V(\bar{e}^2)_{\max}$.2321	5.484×10^4
.001 \bar{e}_{\max}^2											.3453
$V(\bar{e}^2)_{\max}$.2024

$\alpha_p = 0$. In all other cases $m_p = 1$.

TABLE XV. - Continued. EFFECT OF α_p AND α_f ON MEAN SQUARED
ERROR AND ON VARIANCE OF MEAN SQUARED ERROR

(d) $\delta_k = -1, 1, -1, -1, -1, -1, -1, -1; \delta_k = 1, 1, 1, 1, 1, 1, 1, 1;$

$g = 8; h = 4; n_e = 1000; e = 1; \tau = 0$

α_p	α_f										
	1.00	0.75	0.50	0.25	0.10	0.05	0.025	0.01	0.005	0.002	0.001
1.00 \bar{e}_{\max}^2	1.117										
$V(\bar{e}^2)_{\max}$	2.584										
1.00 \bar{e}_{\max}^2	1.120	1.121	1.113	1.162	2.303	4.911	8.114	11.14	12.49	13.16	14.07
$V(\bar{e}^2)_{\max}$	2.589	2.574	2.612	2.611	13.80	31.61	1.799	30.34	2.309	13.54	3.550
.75 \bar{e}_{\max}^2		1.143	1.150	1.175	1.460	2.714	4.936	8.016	10.01	11.51	12.77
$V(\bar{e}^2)_{\max}$		2.704	2.620	2.701	6.330	16.91	32.86	39.31	35.27	26.19	15.95
.50 \bar{e}_{\max}^2			1.433	1.597	1.872	2.451	3.760	6.017	7.679	9.420	11.04
$V(\bar{e}^2)_{\max}$			5.358	7.369	11.01	15.79	26.71	36.79	39.50	35.24	27.99
.25 \bar{e}_{\max}^2				3.391	4.148	4.702	5.364	6.646	7.854	9.166	10.68
$V(\bar{e}^2)_{\max}$				22.55	28.45	32.23	35.60	38.93	39.72	36.22	29.84
.10 \bar{e}_{\max}^2					8.071	8.801	9.345	9.867	10.30	10.91	11.89
$V(\bar{e}^2)_{\max}$					39.50	39.41	38.78	37.36	35.46	31.02	23.96
.05 \bar{e}_{\max}^2						10.79	11.40	11.85	11.98	12.26	12.82
$V(\bar{e}^2)_{\max}$						31.07	28.03	25.66	24.68	22.11	16.90
.025 \bar{e}_{\max}^2							12.42	12.75	12.85	13.01	13.33
$V(\bar{e}^2)_{\max}$							20.63	18.08	17.05	15.46	12.03
.01 \bar{e}_{\max}^2								13.47	13.53	13.64	13.74
$V(\bar{e}^2)_{\max}$								10.47	9.832	8.793	7.712
.005 \bar{e}_{\max}^2									13.71	13.81	13.86
$V(\bar{e}^2)_{\max}$									7.720	6.841	6.289
.002 \bar{e}_{\max}^2										13.94	13.98
$V(\bar{e}^2)_{\max}$										5.236	4.790
.001 \bar{e}_{\max}^2											14.04
$V(\bar{e}^2)_{\max}$											3.867

^a $m_p = 0$. In all other cases $m_p = 1$.

TABLE XV. - Continued. EFFECT OF α_p AND α_f ON MEAN SQUARED
ERROR AND ON VARIANCE OF MEAN SQUARED ERROR

(e) $g = 4$, $h = 0$, $\theta = 0.125$; $\tau = 1.0$; $\delta_k = 1, 1, 1, 1$; $n_e = 1000$

α_p	α_f										
	1.00	0.75	0.50	0.25	0.10	0.05	0.025	0.01	0.005	0.002	0.001
$\alpha_{1.00}$											
\bar{e}_{\max}^2	1.116										
$V(\bar{e}^2)_{\max}$	2.282										
1.00	1.115	1.112	1.101	1.045	0.8551	0.6207	0.4286	0.3335	0.2520	0.2466	0.2195
\bar{e}_{\max}^2											
$V(\bar{e}^2)_{\max}$	2.291	2.311	2.289	2.120	1.653	1.168	.8042	.4406	.2171	.1405	.0460
.75		1.089	1.084	1.056	.9501	.7833	.5991	.4371	.3267	.2927	.2520
\bar{e}_{\max}^2											
$V(\bar{e}^2)_{\max}$		2.275	2.275	2.201	1.807	1.406	1.125	.6759	.4127	.2603	.1137
.50			1.029	.9936	.9239	.8327	.6968	.5251	.3920	.3454	.2889
\bar{e}_{\max}^2											
$V(\bar{e}^2)_{\max}$			2.214	2.142	1.842	1.690	1.358	.9613	.7247	.3803	.2412
.25				.7585	.6967	.6554	.6003	.4966	.3983	.3615	.3060
\bar{e}_{\max}^2											
$V(\bar{e}^2)_{\max}$				1.608	1.542	1.408	1.180	.9287	.8147	.4493	.3128
.10					.4246	.4067	.3899	.3635	.3312	.3153	.2808
\bar{e}_{\max}^2											
$V(\bar{e}^2)_{\max}$.8632	.8405	.7459	.6425	.5259	.3318	.2668
.05						.3276	.3168	.3056	.2947	.2881	.2595
\bar{e}_{\max}^2											
$V(\bar{e}^2)_{\max}$.5294	.5183	.4802	.3888	.2646	.2040
.025							.2847	.2726	.2685	.2616	.2441
\bar{e}_{\max}^2											
$V(\bar{e}^2)_{\max}$.2851	.2593	.2488	.2127	.1353
.01								.2378	.2364	.2352	.2302
\bar{e}_{\max}^2											
$V(\bar{e}^2)_{\max}$.1474	.1423	.1385	.1073
.005									.2353	.2357	.2302
\bar{e}_{\max}^2											
$V(\bar{e}^2)_{\max}$.1404	.1369	.1061
.002										.2229	.2196
\bar{e}_{\max}^2											
$V(\bar{e}^2)_{\max}$.0541	.0460
.001											.2195
\bar{e}_{\max}^2											
$V(\bar{e}^2)_{\max}$.0460

^a $m_p = 0$. In all other cases $m_p = 1$.

TABLE XV. - Continued. EFFECT OF α_p AND α_f ON MEAN SQUARED ERROR AND ON VARIANCE OF MEAN SQUARED ERROR(f) $\delta_k = 1, 1, 1, 1$; $g = 4$; $\theta = 1.0$; $\tau = 1.0$; $h = 0$; $n_e = 1000$

α_p	α_f											
	1.00	0.75	0.50	0.25	0.10	0.05	0.025	0.01	0.005	0.002	0.001	
$\alpha_{1.00}$	\bar{e}_{\max}^2 $V(\bar{e}^2)_{\max}$	1.116 2.282										
1.00	\bar{e}_{\max}^2 $V(\bar{e}^2)_{\max}$	1.117 2.307	1.112 2.327	1.095 2.202	1.060 2.037	1.401 4.132	2.862 1.269 $\times 10$	4.901 1.923 $\times 10$	7.376 1.726 $\times 10$	8.676 1.137 $\times 10$	9.214 8.197	9.952 2.482
.75	\bar{e}_{\max}^2 $V(\bar{e}^2)_{\max}$		1.098 2.342	1.089 2.245	1.064 2.031	1.138 2.508	1.724 6.358	3.092 1.434 $\times 10$	5.116 2.007 $\times 10$	6.532 1.946 $\times 10$	7.436 1.707 $\times 10$	8.507 1.202 $\times 10$
.50	\bar{e}_{\max}^2 $V(\bar{e}^2)_{\max}$			1.148 2.553	1.193 2.750	1.235 3.054	1.455 4.492	2.148 9.166	3.579 1.629 $\times 10$	4.839 1.948 $\times 10$	5.938 1.963 $\times 10$	7.171 1.683 $\times 10$
.25	\bar{e}_{\max}^2 $V(\bar{e}^2)_{\max}$				1.786 5.767	2.246 8.533	2.644 1.114 $\times 10$	3.022 1.303 $\times 10$	3.759 1.646 $\times 10$	4.602 1.874 $\times 10$	5.436 1.949 $\times 10$	6.648 1.860 $\times 10$
.10	\bar{e}_{\max}^2 $V(\bar{e}^2)_{\max}$					3.725 1.515 $\times 10$	4.368 1.723 $\times 10$	4.949 1.823 $\times 10$	5.691 1.943 $\times 10$	6.209 1.925 $\times 10$	6.867 1.843 $\times 10$	7.771 1.553 $\times 10$
.05	\bar{e}_{\max}^2 $V(\bar{e}^2)_{\max}$						5.477 1.835 $\times 10$	6.132 1.795 $\times 10$	6.969 1.725 $\times 10$	7.409 1.624 $\times 10$	7.871 14.94 $\times 10$	8.490 1.222 $\times 10$
.025	\bar{e}_{\max}^2 $V(\bar{e}^2)_{\max}$							6.984 1.696 $\times 10$	7.796 1.463 $\times 10$	8.171 1.304 $\times 10$	8.554 1.139 $\times 10$	8.968 9.240
.01	\bar{e}_{\max}^2 $V(\bar{e}^2)_{\max}$								8.454 1.200 $\times 10$	8.832 9.984	9.167 7.970	9.378 6.669
.005	\bar{e}_{\max}^2 $V(\bar{e}^2)_{\max}$									9.048 8.820	9.345 6.875	9.534 5.564
.002	\bar{e}_{\max}^2 $V(\bar{e}^2)_{\max}$										9.541 5.597	9.686 4.530
.001	\bar{e}_{\max}^2 $V(\bar{e}^2)_{\max}$											9.752 4.114

^a $m_p = 0$. In all other cases $m_p = 1$.

TABLE XV. - Continued. EFFECT OF α_p AND α_f ON MEAN SQUARED ERROR AND ON VARIANCE OF MEAN SQUARED ERROR(g) $g = 4$; $h = 0$; $\theta = 32.0$; $\tau = 1.0$; $\delta_k = 1, 1, 1, 1$; $n_e = 1000$

α_p	α_f										
	1.00	0.75	0.50	0.25	0.10	0.05	0.025	0.01	0.005	0.002	0.001
1.00 \bar{e}_{\max}^2	1.116										
$V(\bar{e}^2)_{\max}$	2.282										
1.00 \bar{e}_{\max}^2	1.894	9.625	41.81	214.9	462.3	115.6 $\times 10$	505.6 $\times 10$	985.5 $\times 10$	100.8 $\times 10^2$	101.0 $\times 10^2$	101.1 $\times 10^2$
$V(\bar{e}^2)_{\max}$	5.341	178.3	121.8 $\times 10$	910.6 $\times 10$	302.4 $\times 10^2$	408.0 $\times 10^3$	100.9 $\times 10^5$	173.8 $\times 10^4$	253.3 $\times 10^3$	736.3 $\times 10^2$	284.4 $\times 10$
.75 \bar{e}_{\max}^2		89.73	134.1	197.1	284.0	384.4	804.5	253.1 $\times 10$	279.2 $\times 10$	293.1 $\times 10$	307.6 $\times 10$
$V(\bar{e}^2)_{\max}$		101.9 $\times 10^2$	135.2 $\times 10^2$	100.8 $\times 10^2$	595.3 $\times 10$	490.3 $\times 10$	135.2 $\times 10^4$	158.9 $\times 10^5$	173.3 $\times 10^5$	170.4 $\times 10^5$	167.4 $\times 10^5$
.50 \bar{e}_{\max}^2			251.0	272.3	290.1	303.2	334.8	728.4	998.2	104.7 $\times 10$	108.0 $\times 10$
$V(\bar{e}^2)_{\max}$			978.7 $\times 10$	597.9 $\times 10$	201.7 $\times 10$	242.9 $\times 10$	387.3 $\times 10^2$	181.6 $\times 10^4$	377.5 $\times 10^4$	392.2 $\times 10^4$	406.9 $\times 10^4$
.25 \bar{e}_{\max}^2				298.1	298.4	299.3	309.2	528.8	642.1	667.9	679.8
$V(\bar{e}^2)_{\max}$				104.6 $\times 10$	905.9	632.5	340.3 $\times 10$	261.4 $\times 10^2$	108.0 $\times 10^3$	294.8 $\times 10^3$	448.3 $\times 10^3$
.10 \bar{e}_{\max}^2					300.2	300.2	309.4	527.7	628.1	636.5	642.4
$V(\bar{e}^2)_{\max}$					458.3	441.7	336.3 $\times 10$	249.5 $\times 10^2$	346.2 $\times 10$	124.6 $\times 10^2$	885.5 $\times 10^2$
.05 \bar{e}_{\max}^2						300.4	309.4	527.7	627.1	633.0	633.0
$V(\bar{e}^2)_{\max}$						370.0	336.3 $\times 10$	249.5 $\times 10^2$	258.4 $\times 10$	644.4	644.4
.025 \bar{e}_{\max}^2							309.4	527.7	627.1	636.0	662.0
$V(\bar{e}^2)_{\max}$							336.3 $\times 10$	249.5 $\times 10^2$	258.4 $\times 10$	377.0 $\times 10$	298.0 $\times 10^2$
.01 \bar{e}_{\max}^2								527.7	627.1	746.6	132.4 $\times 10$
$V(\bar{e}^2)_{\max}$								249.5 $\times 10^2$	258.4 $\times 10$	105.8 $\times 10^3$	237.8 $\times 10^3$
.005 \bar{e}_{\max}^2									627.1	793.7	161.8 $\times 10$
$V(\bar{e}^2)_{\max}$									258.4 $\times 10$	142.7 $\times 10^3$	495.6 $\times 10^2$
.002 \bar{e}_{\max}^2										215.6 $\times 10$	299.4 $\times 10$
$V(\bar{e}^2)_{\max}$										121.4 $\times 10^5$	974.7 $\times 10^4$
.001 \bar{e}_{\max}^2											982.3 $\times 10$
$V(\bar{e}^2)_{\max}$											262.3 $\times 10^4$

^a $m_p = 0$. In all other cases $m_p = 1$.

TABLE XV. - Continued. EFFECT OF α_p AND α_f ON MEAN SQUARED ERROR AND ON VARIANCE OF MEAN SQUARED ERROR(g) $g = 4$; $h = 0$; $\theta = 32.0$; $\tau = 1.0$; $\delta_k = 1, 1, 1, 1$; $n_e = 1000$

α_p	α_f										
	1.00	0.75	0.50	0.25	0.10	0.05	0.025	0.01	0.005	0.002	0.001
1.00 \bar{e}_{\max}^2	1.116										
$V(\bar{e}^2)_{\max}$	2.282										
1.00 \bar{e}_{\max}^2	1.894	9.625	41.81	214.9	462.3	115.6 $\times 10$	505.6 $\times 10$	985.5 $\times 10$	100.8 $\times 10^2$	101.0 $\times 10^2$	101.1 $\times 10^2$
$V(\bar{e}^2)_{\max}$	5.341	178.3	121.8 $\times 10$	910.6 $\times 10$	302.4 $\times 10^2$	408.0 $\times 10^3$	100.9 $\times 10^5$	173.8 $\times 10^4$	253.3 $\times 10^3$	736.3 $\times 10^2$	284.4 $\times 10$
.75 \bar{e}_{\max}^2		89.73	134.1	197.1	284.0	384.4	804.5	253.1 $\times 10$	279.2 $\times 10$	293.1 $\times 10$	307.6 $\times 10$
$V(\bar{e}^2)_{\max}$		101.9 $\times 10^2$	135.2 $\times 10^2$	100.8 $\times 10^2$	595.3 $\times 10$	490.3 $\times 10$	135.2 $\times 10^4$	158.9 $\times 10^5$	173.3 $\times 10^5$	170.4 $\times 10^5$	167.4 $\times 10^5$
.50 \bar{e}_{\max}^2			251.0	272.3	290.1	303.2	334.8	728.4	998.2	104.7 $\times 10$	108.0 $\times 10$
$V(\bar{e}^2)_{\max}$			978.7 $\times 10$	597.9 $\times 10$	201.7 $\times 10$	242.9 $\times 10$	387.3 $\times 10^2$	181.6 $\times 10^4$	377.5 $\times 10^4$	392.2 $\times 10^4$	406.9 $\times 10^4$
.25 \bar{e}_{\max}^2				298.1	298.4	299.3	309.2	528.8	642.1	667.9	679.8
$V(\bar{e}^2)_{\max}$				104.6 $\times 10$	905.9	632.5	340.3 $\times 10$	261.4 $\times 10^2$	108.0 $\times 10^3$	294.8 $\times 10^3$	448.3 $\times 10^3$
.10 \bar{e}_{\max}^2					300.2	300.2	309.4	527.7	628.1	636.5	642.4
$V(\bar{e}^2)_{\max}$					458.3	441.7	336.3 $\times 10$	249.5 $\times 10^2$	346.2 $\times 10$	124.6 $\times 10^2$	885.5 $\times 10^2$
.05 \bar{e}_{\max}^2						300.4	309.4	527.7	627.1	633.0	633.0
$V(\bar{e}^2)_{\max}$						370.0	336.3 $\times 10$	249.5 $\times 10^2$	258.4 $\times 10$	644.4	644.4
.025 \bar{e}_{\max}^2							309.4	527.7	627.1	636.0	662.0
$V(\bar{e}^2)_{\max}$							336.3 $\times 10$	249.5 $\times 10^2$	258.4 $\times 10$	377.0 $\times 10$	298.0 $\times 10^2$
.01 \bar{e}_{\max}^2								527.7	627.1	746.6	132.4 $\times 10$
$V(\bar{e}^2)_{\max}$								249.5 $\times 10^2$	258.4 $\times 10$	105.8 $\times 10^3$	237.8 $\times 10^3$
.005 \bar{e}_{\max}^2									627.1	793.7	161.8 $\times 10$
$V(\bar{e}^2)_{\max}$									258.4 $\times 10$	142.7 $\times 10^3$	495.6 $\times 10^2$
.002 \bar{e}_{\max}^2										215.6 $\times 10$	299.4 $\times 10$
$V(\bar{e}^2)_{\max}$										121.4 $\times 10^5$	974.7 $\times 10^4$
.001 \bar{e}_{\max}^2											982.3 $\times 10$
$V(\bar{e}^2)_{\max}$											262.3 $\times 10^4$

^a $m_p = 0$. In all other cases $m_p = 1$.

TABLE XV. - Concluded. EFFECT OF α_p AND α_f ON MEAN SQUARED

ERROR AND ON VARIANCE OF MEAN SQUARED ERROR

(h) $g = 8$; $h = 4$; $\theta = 0.125$; $\tau = 1.0$; $\delta_k = 1, 1, 1, 1, \dots$; $n_e = 1000$

α_p		α_f										
		1.00	0.75	0.50	0.25	0.10	0.05	0.025	0.01	0.005	0.002	0.001
1.00	\bar{e}_{\max}^2	1.121										
	$V(\bar{e}^2)_{\max}$	2.569										
1.00	\bar{e}_{\max}^2	1.122	1.122	1.113	1.068	0.9152	0.7405	0.5716	0.4900	0.4452	0.4223	0.4081
	$V(\bar{e}^2)_{\max}$	2.559	2.568	2.551	2.401	2.003	1.579	.9782	.7405	.4034	.3373	.0924
.75	\bar{e}_{\max}^2		1.108	1.105	1.084	.9826	.8524	.6830	.5581	.4771	.4476	.4295
	$V(\bar{e}^2)_{\max}$		2.556	2.517	2.463	2.160	1.752	1.262	.8933	.5934	.4207	.1618
.50	\bar{e}_{\max}^2			1.042	1.028	.9848	.8971	.7506	.6397	.5226	.4836	.4533
	$V(\bar{e}^2)_{\max}$			2.443	2.338	2.161	1.971	1.508	1.208	.8133	.5711	.3190
.25	\bar{e}_{\max}^2				.8118	.7891	.7508	.6932	.6272	.5488	.5007	.4707
	$V(\bar{e}^2)_{\max}$				1.991	1.886	1.796	1.433	1.211	.9575	.7156	.4482
.10	\bar{e}_{\max}^2					.5684	.5603	.5262	.5105	.4826	.4598	.4491
	$V(\bar{e}^2)_{\max}$					1.150	1.082	1.031	1.058	.8848	.6804	.4114
.05	\bar{e}_{\max}^2						.4745	.4583	.4517	.4425	.4278	.4273
	$V(\bar{e}^2)_{\max}$.7093	.6980	.6862	.5831	.5700	.2981
.025	\bar{e}_{\max}^2							.4422	.4420	.4428	.4327	.4265
	$V(\bar{e}^2)_{\max}$.4933	.4947	.4475	.4851	.2503
.01	\bar{e}_{\max}^2								.4227	.4242	.4207	.4185
	$V(\bar{e}^2)_{\max}$.3530	.3454	.3384	.1419
.005	\bar{e}_{\max}^2									.4160	.4158	.4137
	$V(\bar{e}^2)_{\max}$.1957	.1965	.1302
.002	\bar{e}_{\max}^2										.4091	.4079
	$V(\bar{e}^2)_{\max}$.1787	.0925
.001	\bar{e}_{\max}^2											.4078
	$V(\bar{e}^2)_{\max}$.0925

^a $m_p = 0$. In all other cases $m_p = 1$.

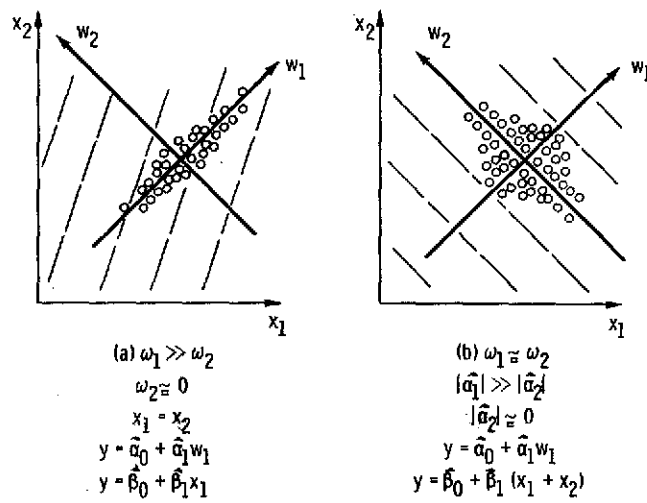


Figure 1. - Principal components regression.

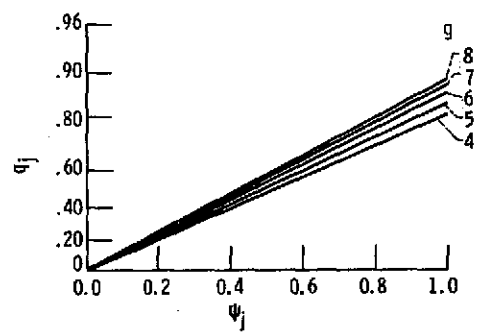
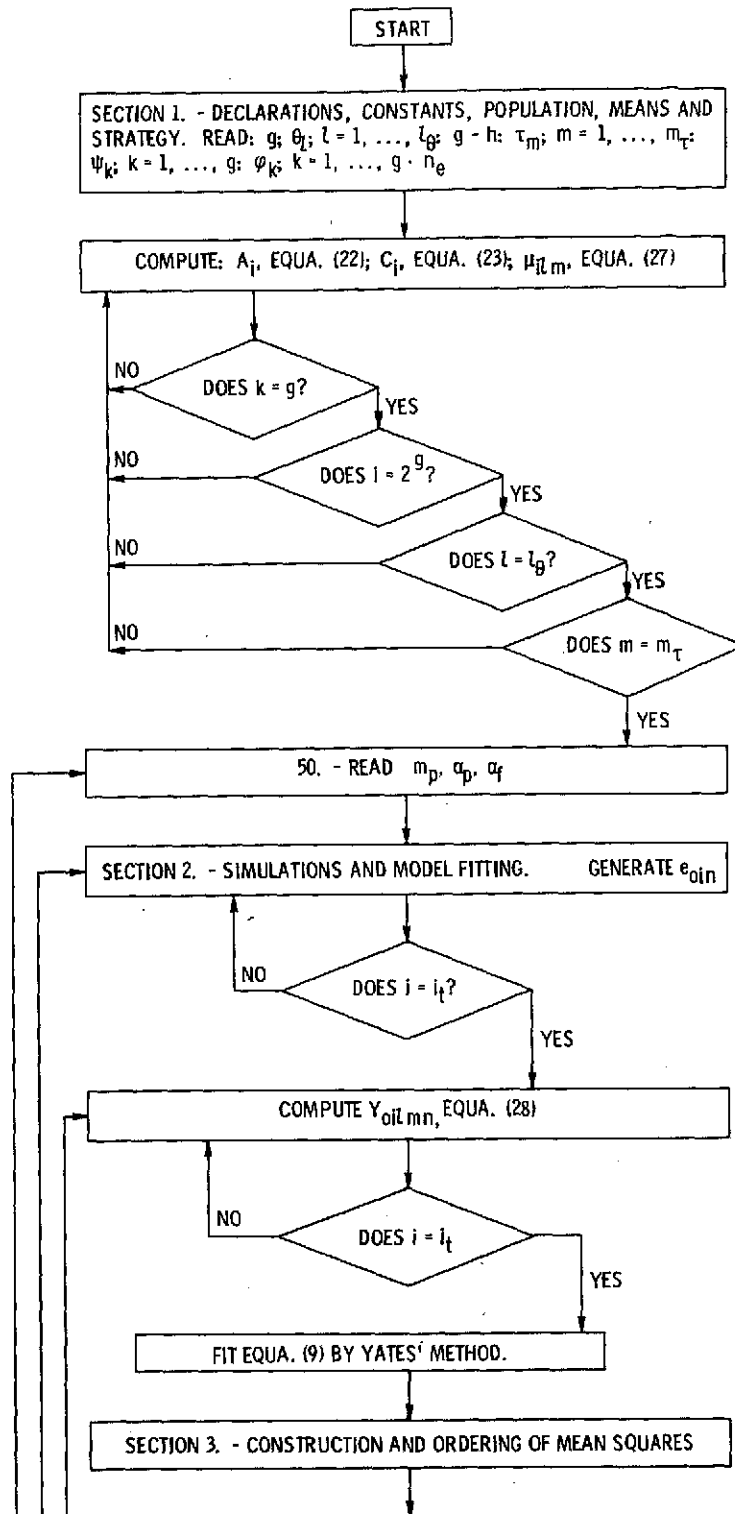
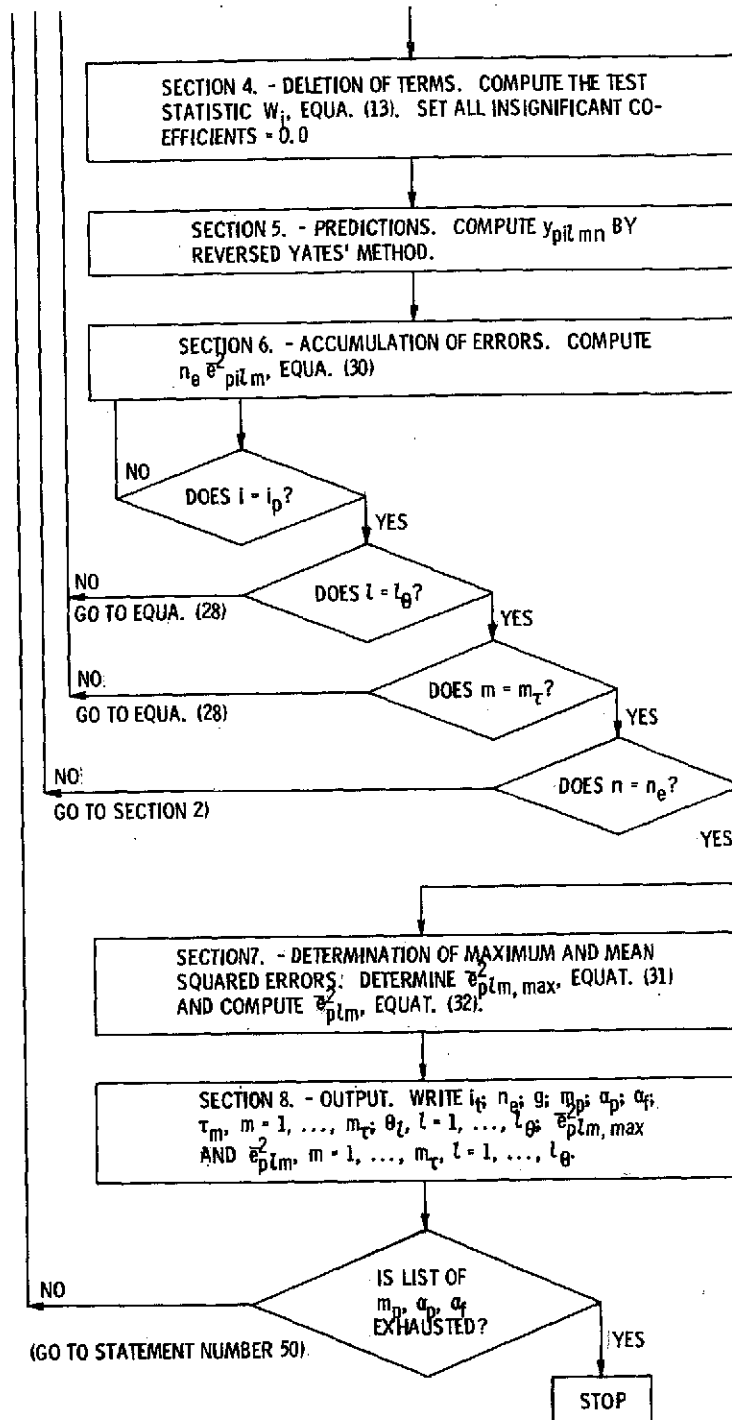


Figure 2. - Half normal plot.



(a) SECTIONS 1, 2, AND 3.

Figure 3. - Flow chart for POOLs.



(b) SECTIONS 4, 5, 6, 7, AND 8.

Figure 3. - Concluded.

ABSTRACT

The investigation consisted of Monte Carlo studies using population models intended to represent response surface applications. Simulated experiments were generated by adding pseudo random normally distributed errors to population values to generate "observations." Model equations were fitted to the "observations" and the decision procedure was used to delete terms. Comparison of values predicted by the reduced models with the true population values enabled the identification of deletion strategies that are approximately optimal for minimizing prediction errors.

KEY WORDS

Chain pooling; Subset regression; Statistical analysis; Regression analysis; Mathematical models; Model selection; Principal components regression; Data reduction; Statistical tests; Statistical decision theory